Connection Science
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713411269

Using unsupervised incremental learning to cope with gradual concept drift

David Hadas[a]; Galit Yovel[b]; Nathan Intrator[a]
[a] School of Computer Science, Tel-Aviv University, Ramat Aviv, Israel [b] Department of Psychology, Tel-Aviv University, Ramat Aviv, Israel

Online publication date: 13 May 2011

To cite this Article Hadas, David , Yovel, Galit and Intrator, Nathan(2011) 'Using unsupervised incremental learning to cope with gradual concept drift', Connection Science, 23: 1, 65 — 83
To link to this Article: DOI: 10.1080/09540091.2011.575929
URL: http://dx.doi.org/10.1080/09540091.2011.575929

Connection Science
Vol. 23, No. 1, March 2011, 65–83

Taylor & Francis
Taylor & Francis Group

Using unsupervised incremental learning to cope with gradual concept drift

David Hadas[a]*, Galit Yovel[b] and Nathan Intrator[a]

[a]*School of Computer Science, Tel-Aviv University, Ramat Aviv, Israel;* [b]*Department of Psychology, Tel-Aviv University, Ramat Aviv, Israel*

(*Received 1 November 2010; final version received 24 March 2011*)

Current computational approach to incremental learning requires a constant stream of labelled data to cope with gradual environmental changes known as concept drift. This paper examines a case where labelled data are unavailable. Inspired by the performance of the human visual system, capable of adjusting its concepts using unlabelled stimuli, we introduce a variant to an unsupervised competitive learning algorithm known as the Leader Follower (LF). This variant can adjust pre-learned concepts to environmental changes using unlabelled data samples. We motivate the needed change in the existing LF algorithm and compare between two variants to enable the accumulation of environmental changes when facing unbalanced sample ratio.

Keywords: unsupervised competitive learning; unsupervised learning; incremental learning; online learning; leader follower

1. Introduction

Gradual changes in the environment may reduce a machine classification accuracy (Gao, Fan, and Han 2007). Initially, the machine may learn a set of concepts. Using such learning, when the machine is given an unlabelled sample from the environment, it may be able to predict whether the sample belongs to the concepts learned. Following environmental changes, samples once belonging to one concept may later belong to another concept. The accumulation of such changes would turn a stationary machine learner useless. This paper discusses the use of unsupervised learning to maintain the accuracy of a machine learner during gradual environmental changes and presents a biologically inspired algorithm to adapt the learner to the accumulation of such changes.

A dynamic learning machine that is able to adapt its pre-learned concepts during the classification process may be able to sustain environmental changes and remain reasonably accurate (Helmbold and Long 1994). Under certain problem areas, true labels are either hard to achieve or not available in a timely manner; see, for example, Kelly, Hand, and Adams (1999) describing a bank loan application where true labels become available 24 months after a loan was granted.

*Corresponding author. Email: david.hadas@gmail.com

ISSN 0954-0091 print/ISSN 1360-0494 online

DOI: 10.1080/09540091.2011.575929
http://www.informaworld.com

Under such circumstances, Kuncheva (2008) suggested to use unlabelled samples to detect concept drift. Here, we discuss the use of unlabelled samples to adapt the machine pre-learned concepts and maintain its accuracy.

A distinction can be made between an abrupt environmental change and a gradual one. In the case of an abrupt change, the environment after the change is not well correlated to the original environment prior to the change; labelling the new environment and training a new learner is one approach for coping with an abrupt change of the environment (Kuncheva 2004). Yet, in many cases, the changes are gradual such that at every juncture, the environment can be well correlated to a previous juncture. Still, the accumulation of gradual changes may alter the environment dramatically over time. We study the use of unlabelled samples and the correlation between junctures to tune pre-learned concepts and follow gradual environmental changes.

1.1. *State of the art*

In machine learning, learning under a concept drift is usually considered in the context of supervised learning; see, for example, performance comparison by Klinkenberg and Renz (1998), method classification by Maloof and Michalski (2004), and a review by Zliobaite (2009). Unsupervised learning methods are sometimes suggested for detecting concept drift (for example Kuncheva 2008) and is also mentioned in the context of sensor drift (for example, Natale, Fabrizio, and D'Amico 1995).

On-line supervised learners may use a stream of new labelled samples to gradually change with time and adapt to concept drift (Hulten, Spencer, and Domingos 2001; Tsymbal 2004). Some supervised learners use instance selection methods in which a moving window determines the instances used by the model; see, for example, FLORA by Widmer and Kubat (1996), Time-Windowed Forgetting by Salganicoff (1997), OnlineTree2 by Nunez, Fidalgo, and Morales (2007) and a method based on Support Vector Machines by Klinkenberg and Joachims (2000). Other supervised learners use instance weighting methods in which instances are weighted according to their age and competence with regard to the current concept (see, for example, Klinkenberg 2004). Yet another approach is to use an ensemble of classifiers where new labels are used to either train new classifiers of the ensemble or to weight existing ones (Kuncheva 2004; Elwell and Polikar 2009). Some supervised learners are suitable for classifying imbalanced class distributions (Chen and He 2009; Ditzler, Muhlbaier, and Polikar 2010). All such supervised on-line learners depend on timely availability of labelled samples. Adapting to concept drift without such dependency would require us to explore unsupervised learning methods.

Reviewing unsupervised learning literature, one may divide the available algorithms by their concept stability; some unsupervised learning algorithms use a global criteria optimisation and as a result, a new sample may cause major changes to the pre-learned concepts; in the context of on-line unsupervised learning, such algorithms are said to be unstable Duda, Hart, and Stork 2000). As an example, adding a single sample to the K-Means algorithm may change all resulting clusters; a behaviour which may be deemed unacceptable when tuning concepts previously learned from labelled samples. The requirement for concept structure stability is known as the Stability-Plasticity Dilemma (Carpenter and Grossberg 1988).

Incremental unsupervised learning algorithms avoid using a global criteria optimisation and maintain concept structure stability; given a new sample, an incremental unsupervised learner would adjust only concepts in the proximity of the sample; see Duda et al. (2000) but also Unsupervised Competitive Learning by Kosko (1991) and Kong and Kosko (1991) and work by Tsypkin (1973). As new unlabelled samples arrive, unsupervised algorithms such as CLASSIT (Gennari, Langley, and Fisher 1989), ART Carpenter and Grossberg (1988), M-SOM-ART Zehraoui and Bennani (2004) and the Leader Follower (LF) algorithm (Duda et al. 2000), continue to incrementally fine-tune the concepts while maintaining a stable concept structure. A similar incremental learning approach is also used by Adaptive Vector Quantization techniques. (See, for example, Gersho and Yano 1985 describing a progressive codevector replacement technique.)

The LF algorithm may be considered one of the simplest unsupervised incremental learning algorithms in active use. Since its introduction, LF has been used in various applications requiring incremental learning of clusters from unlabelled data. Some examples include: Offline Speaker Clustering, where LF has shown promising results compared with existing offline speaker clustering, while running much more efficiently (Liu and Kubala 2004); project ARGUS, where LF is part of a novelty detection mechanism (Carbonell et al. 2006); a distributed version of LF was suggested to enable the formation of Mobile Ad Hoc Networks (Garg and Shyamasundar 2004); in Pei, Xu, Zheng, and Bin (2005), an LF clustering method is used for continuous classification of hand motor imagery tasks; (Fan et al. 2007) uses LF as part of a Film cast indexing method.

Incremental unsupervised learning algorithms have not been specifically adopted to cope with concept drift or to tune concepts pre-learned by a supervised learner.

1.2. *The biological visual system and concept drift*

Adaptation of the biological visual system to concept drift was recently studied during physiological and psychophysical experimentations (Wallis and Bulthoff 2001; Leutgeb et al. 2005; Preminger, Sagi, and Tsodyks 2007; Hadas, Intrator, and Yovel 2010).

When facing environmental changes, such as familiar objects that change with time, the visual system was shown to use the following strategy: (1) After an initial phase in which concepts were learned with labels, the visual system is capable of drifting pre-learned concepts to follow environmental changes; (2) Such adaptation is demonstrated with unlabelled samples; (3) Unlabelled samples affect future classifications of resembling stimuli; (4) While using such unsupervised incremental learning, the visual system can radically drift pre-learned concepts and accumulate environmental changes; (5) The visual system is able to handle concept drift even if the drift is towards a region in the concept space that was previously learned as belonging to a different concept. The concept can be drifted to completely overlap such a region. This was demonstrated with a balanced sample distribution psychophysical experiment in which both concepts were presented (Preminger et al. 2007) and more recenly with an unbalanced sample distribution psychophysical experiment in which only the drifting concept was presented (Hadas et al. 2010).

1.3. *The machine learner challenge*

The biological reports presented above inspired us to revisit unsupervised learning algorithms, in search of mechanisms that will maintain the accuracy of machine learners without depending on timely labelled samples. Motivated by such reports, it would be desirable to further investigate incremental unsupervised machine learning algorithms and modify them to posses the following features: (A) Given a machine learner initially trained by a set of labelled samples, use recent unlabelled samples to adapt the machine to concept drift; (B) Support the accumulation of drift using unlabelled samples; (C) Support drifting of a concept towards regions that were previously learned as belonging to different concepts using unlabelled samples.

In this paper, we introduce an unsupervised 'Concept Follower' algorithm that is capable of tuning concepts, which were pre-learned by a supervised learner. We adapted the LF algorithm (Duda et al. 2000) to cope with problem areas in which true labels are either hard to achieve or not available in a timely manner, yet concept drift is expected. In such problem areas, the use of existing supervised methods for coping with concept drift is inadequate (Kuncheva 2004). The resulting algorithms offer the ability to adjust pre-learned concepts when facing gradually accumulating environmental changes. It is shown that by using unlabelled samples, a Concept Follower based on LF may handle concept drift even when the drifted concept completely overlaps a region previously learned as the centre of a different concept.

2. Methodology

In this section, unsupervised Concept Followers (CFs) are presented and analyzed. The strategy used includes an initial supervised stage. During the supervised stage, concept centres are learned from a multi-dimensional environment using labelled samples. Once concepts were learned, the presented CFs use unlabelled samples to accommodate the concept centres to environmental changes. Such bootstrapping is achieved through incremental learning and is targeted for coping with gradual accumulation of environmental changes. When the changes are not gradual but abrupt, the CFs detect the abrupt change, to allow repeating the supervised stage.

The traditional LF algorithm (Duda et al. 2000) is presented first. Than, a Concept Follower algorithm (CF1) is derived by adjusting LF to a new function of drifting pre-learned concepts and following environmental changes. Last, a novel CF with Unlearning algorithm (CF2) is presented. CF2 adds an unlearning mechanism side by side to the learning mechanism of CF1. The suggested unlearning mechanism is designed to increase the information induced by each unlabelled sample.

2.1. *The traditional Leader Follower algorithm*

LF is an unsupervised competitive learning algorithm. After each classification, the algorithm uses the new sample and its assigned label to adjust the concept, which it was classified as belonging to. If the sample was not classified as belonging to any concept, then the sample is considered as a concept of its own. Consequently, the LF can be applied when the number of classes is unknown.

In the original (Duda et al. 2000) algorithm, adjusting the concept towards the sample was achieved by adding the new sample to the concept and normalising the resulting concepts at each step. A generalisation of the (Duda et al. 2000) algorithm, which allowed any form of concept update towards the sample was presented by Garg and Shyamasundar (2004). This variant included learning without concept and sample normalisation.

The (Garg and Shyamasundar 2004) variant of the traditional Leader Follower algorithm is presented in Algorithm 1. The algorithm initialises the first concept (w_1) using the first sample and collects additional concepts (w_i for $i > 1$) as it proceeds. Samples not in the proximity of any existing concept are used as the basis for new concepts. Samples at the proximity of a learned concept are considered as belonging to that concept. The proximity measure is controlled by a threshold parameter (θ).

Once a sample is classified as belonging to a concept, learning takes place and the concept is adjusted towards the new sample. Adjusting the concept towards the sample can be considered as a method of ensuring that similar samples would be classified the same as the current one.

The LF mechnism can be adapted to serve as the basis of a CF. Several modifications are required to the LF mechanisms presented above: (1) LF adapts self-learned concepts while the CF should adapt concepts initially learned by a supervised learner. (2) LF may change the number of learned concepts during its operation while a CF should not change the number of concepts. Under the CF strategy suggested here, conceptualisation of the environment is a prerogative of the supervised learner. (3) LF has no mechanism to monitor its accuracy while a CF should monitor its accuracy and identify when supervised learning should be reused.

CF1 and CF2, presented next, include such modifications.

Algorithm 1 The traditional Leader Follower algorithm: LF()

- Initialise θ to the distance threshold between a sample and a concept

$s \leftarrow$ New sample
$N \leftarrow 1$
$w_N \leftarrow s$
loop
    $s \leftarrow$ New sample
    $j \leftarrow \underset{i}{\operatorname{argmin}} \|s - w_i\|$ for any $i = 1..N$
    if $\|s - w_j\| < \theta$ **then**
        Add s to w_j and update w_j towards s
    else
        $N \leftarrow N + 1$
        $w_N \leftarrow s$
    end if
end loop

2.2. *The CF1 algorithm*

A first adaptation of LF to drift concepts in-line with gradual environmental changes is presented in Algorithm 2.

The presented solution framework uses a supervisor to initially learn and label a set of concepts (w_i) and to predict the sample error rate (ϵ_0). Then a CF1 algorithm, is used to drift the pre-learned concepts (w_i). The concepts originally learned using older labelled samples are later adjusted based on newer unlabelled samples. This allows accommodating the concepts to changes in the environment.

The CF1 algorithm is derived from LF. Unlike LF, new concepts are not learned by CF1. Instead, samples that are not in the proximity of any concept are ignored. We use the ratio of such samples in the sample set as an indicator of the algorithm health. Once the error ratio crosses some boundary ($\epsilon_{\max}$), the algorithm is no longer considered to be tuned to the environment. In such a case, it may be appropriate to use a supervised learner once again. The error rate boundary value ($\epsilon_{\max}$) should be set higher than the error rate predicted during the initial supervised learning phase (i.e., $\epsilon_{\max} > \epsilon_0$).

The above CF1 variant selects and adjusts the concept closest to the sample (the one with the minimal distance to the sample). This is similar to the method presented in the LF algorithm. First, the distance of all concepts is calculated and the concept with the minimal distance to the sample is identified. Then the distance of that concept is compared with a predefined threshold parameter (θ). If the distance is smaller than the threshold, then the concept is considered a match and the concept is slightly shifted towards the classified sample to help correct to concept drift. The shift is controlled by a learning rate parameter (η). Adjusting the concept towards the sample can be considered as a method of increasing the likelihood that similar samples would be classified the same as the current one.

The CF1 algorithm may radically drift pre-learned concepts and accumulate environmental changes. Yet, CF1 relies on the availability of sufficient samples from each pre-learned concept. Some problem areas may introduce unbalanced sample ratio between concepts. In such problem areas, CF1 may fail to drift a first concept towards a region previously learned as belonging to a second concept. This may occur, for example, when the sample ratio of the second concept is

Algorithm 2 The Concept Follower algorithm: CF1()

- Initialise $N \geq 1$ to the number of concepts learned using a supervised learner
- Initialise $\{w_i | i = 1..N\}$ to the concepts learned using a supervised learner
- Initialise θ to the distance threshold between a sample and a concept
- Initialise η to the concept learning rate (e.g. $\eta = 0.1$)
- Initialise $\epsilon_{\max}$ to the maximal ratio of allowed errors (e.g. $\epsilon_{\max} = 2 * \epsilon_0$)
- Initialise T_{window} in which to evaluate the error ratio (e.g. $T_{window} = 30/\epsilon_{\max}$)

$T \leftarrow 0$
$E \leftarrow 0$
loop
  $T \leftarrow T + 1$
  $s \leftarrow$ New sample
  $j \leftarrow \underset{i}{\operatorname{argmin}} \|s - w_i\|$ for any $i = 1..N$
  if $\|s - w_j\| < \theta$ **then**
    $w_j \leftarrow w_j + \eta \cdot (s - w_j)$
  else
    $E \leftarrow E + 1$
  end if
  if $T > T_{window}$ **then**
    if $E/T > \epsilon_{\max}$ **then**
      Terminate and initiate new supervised learning
    end if
    $T \leftarrow 0$
    $E \leftarrow 0$
  end if
end loop

too low. We suggest here next that problem areas with unbalanced sample ratio between concepts may consider adjusting all concepts in the proximity of the sample rather than only the concept closest to the sample. Such a modification, may help adjust concepts with low sample ratio.

2.3. *The Concept Follower with unlearning algorithm (CF2)*

A second adaptation of LF is presented in Algorithm 3. The CF2 variant increases the information learned from each sample as compared to LF and CF1. CF2 adjusts all concepts in the proximity of the sample while LF and CF1 adjust only the concept closest to the sample. This allows accommodating concepts to changes in the environment in problem areas with unbalanced sample ratio between concepts. In such problem areas, CF2 can drift a first concept towards a region previously learned as belonging to a second concept even when the sample ratio of the second concept is low.

Adjusting all concepts in the proximity of the sample is done by introducing an unlearning mechanism to the learning mechanism used by LF and CF1. The suggested unlearning mechanism is symmetrical to the learning one. Learning and unlearning are used only when concepts are at the proximity of the sample. Consequently, the algorithm avoids unnecessary adaptation of previously learned concepts, and maintains the stability of the concept structure.

Algorithm 3 Concept Follower with Unlearning algorithm: CF2()

- Initialise $N \geq 1$ to the number of concepts learned using a supervised learner
- Initialise $\{w_i | i = 1..N\}$ to the concepts learned using a supervised learner
- Initialise θ to the distance threshold between a sample and a concept
- Initialise η to the concept learning rate (e.g. $\eta = 0.1$)
- Initialise δ to the concept unlearning rate (e.g. $\delta = 0.05$)
- Initialise $\epsilon_{\max}$ to the maximal ratio of allowed errors (e.g. $\epsilon_{\max} = 2 * \epsilon_0$)
- Initialise T_{window} in which to evaluate the error ratio (e.g. $T_{window} = 30/\epsilon_{\max}$)

$T \leftarrow 0$
$E \leftarrow 0$
loop
    $T \leftarrow T + 1$
    $s \leftarrow$ New sample
    $j \leftarrow \operatorname*{argmin}_i \|s - w_i\|$ for any $i = 1..N$
    if $\|s - w_j\| < \theta$ **then**
        $w_j \leftarrow w_j + \eta \cdot (s - w_j)$
        $w_i \leftarrow w_i - \delta \cdot (s - w_i)$ for all $i = 1..N; i \neq j; \|s - w_i\| < \theta$
    else
        $E \leftarrow E + 1$
    end if
    if $T > T_{window}$ **then**
        if $E/T > \epsilon_{\max}$ **then**
            Terminate and initiate new supervised learning
        end if
        $T \leftarrow 0$
        $E \leftarrow 0$
    end if
end loop

When a new sample is processed, the distances of all concepts to the sample are calculated and compared with a predefined threshold parameter (θ). All concepts with below threshold distance are declared competitors. When there are no competitors the sample is ignored. The ratio of such samples in the sample set is used as an indicator of the algorithm health. Like in CF1, once the error ratio crosses some boundary $(\epsilon_{\max})$, the algorithm is no longer considered to be tuned to the environment. When there is only one concept which is a competitor, the concept would perform learning as in the CF1 algorithm. When there are multiple concepts which are competitors, CF2 would adapt all competing concepts.

When multiple concepts have below threshold distance, a 'winner takes all' procedure is used. The concept with the smallest distance to the sample would perform learning as in CF1. As a result of such learning, the concept is shifted towards the sample. The shift is controlled by a learning rate parameter (η). All other competing concepts with distance below the threshold would perform unlearning. As a result of unlearning, these concepts are shifted away from the sample. The shift is controlled by an unlearning rate parameter (δ). Adjusting the winning concept towards the sample can be considered as a method of increasing the likelihood that similar samples would be classified the same as the current one. Adjusting the other competing concepts away from the sample can be considered as a method of decreasing the likelihood that similar samples would be classified as one of these competing concepts.

We have introduced two algorithms which are capable of adjusting without labels a set of concepts to environmental changes. Both algorithms allow initial learning using a supervised learner and later adaptation to concept drift without labels. In CF1, we suggested to adjust only the concept closest to the sample. In order to support problems areas with unbalanced sample ratio between concepts we introduced CF2 and suggested that in some problem areas it may be beneficial to adjust all concepts in the proximity of the sample.

As a reference, we use a simple supervised incremental learner. Our incremental learner maintains a sliding window with the 10 latest labelled samples per each concept. The Concept centre is calculated using the 10 labels.

3. Experimental setup

In order to evaluate the behaviour of the CF1 and CF2 algorithms, we used samples taken from a morphing sequence between two face image exemplars. One hundred and one samples were taken from the morphing sequence representing a gradual change from the first face image to the second. The morphing sequence used here was derived from the one used in Hadas et al. (2010). Two front face images were handpicked from a 100-face nottingham_scans image set (downloaded from http://pics.psych.stir.ac.uk/index.html). Each of the faces was prepared by blacking out the background, hair and ears. The faces were used for the preparation of a morphing sequence that included 99 intermediate morphed images. The sequence was prepared using Sqirlz Morph version 1.2e. As a result of these preparation steps, we had 101 ordered face images representing a smooth and gradual morphing sequence between the two original face images. The stimuli was later adapted to suit a machine learner.

As part of the preparations to the computational analysis, the 101 images from the morphing sequence were centred using the nose and eyes and a low resolution 19 pixel wide and 29 pixel height image was extracted to include the information of the eyes, eyebrows and mouth (Figure 1). Each of the 101 low-resolution images was represented using a 551 dimensional vector to depict the $(29 * 19 =)551$ pixels of the low-resolution image. Each dimension was in the range of 0 (representing a black pixel) and 255 (representing a white pixel). The experiment uses the resulting vectors as 101 possible samples from the morphing sequence. The 101 possible samples were numbered in the range of 100–200 such that the morphing sequence starts with sample 100 representing the first original face and gradually continues towards sample 200 representing the second original face.

The unsupervised incremental learning algorithms, CF1 and CF2, were examined using different learning, unlearning, and threshold parameters. The performance of the unsupervised learner was evaluated under the assumption that the concepts were pre-learned using a flawless supervised learner. In our evaluation, instead of using a supervised learner, the concept centres were programmatically set to the two face image exemplars. The unsupervised algorithm concepts were

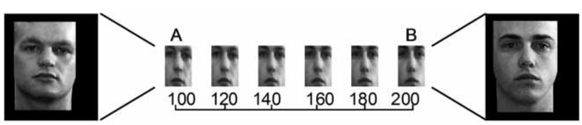

Figure 1. Two face morphing sequence. Two real face images are picked. Ninety nine additional morphed images are created using a morphing utility. Low-resolution abstracts from the 101 images are then used as possible samples and are numbered 100–200.

initialised to include two concepts – Concept 'A' which was set to sample 100 and concept 'B' which was set to sample 200.

Two presentation protocols were used; a sequential presentation and a random one. In the sequential presentation protocol, the machine learner processed the 101 possible samples in sequence starting from sample 100 and ending with sample 200. The sequence was presented once. In the random presentation protocol, 5000 samples were selected uniformly in random from the 101 possible samples of the morphing sequence.

In order to evaluate the progress made by the machine learner during the incremental learning process, we assessed the learner state after each sample (Figure 2). The assessment process after each sample is shown in Table 1. The assessment included calculating the proximity of Concept 'A' and Concept 'B' to each of the 101 possible samples. The sample that was closest to Concept 'A', following each sample learned was named the concept centre of 'A'. The sample that was closest to Concept 'B', following each sample learned was named the concept centre of 'B'.

A second evaluation of CF2 is also presented here. Eight images of shapes were used, each 40 pixel wide and 40 pixel height. The background was set to light gray RGB(180,180,180) and the shapes were drawn using dark gray RGB(70,70,70). The shapes included isosceles triangle, a right-angled triangle, a circle, a half-circle, an ellipse, a square, a rectangle, and an X shape.

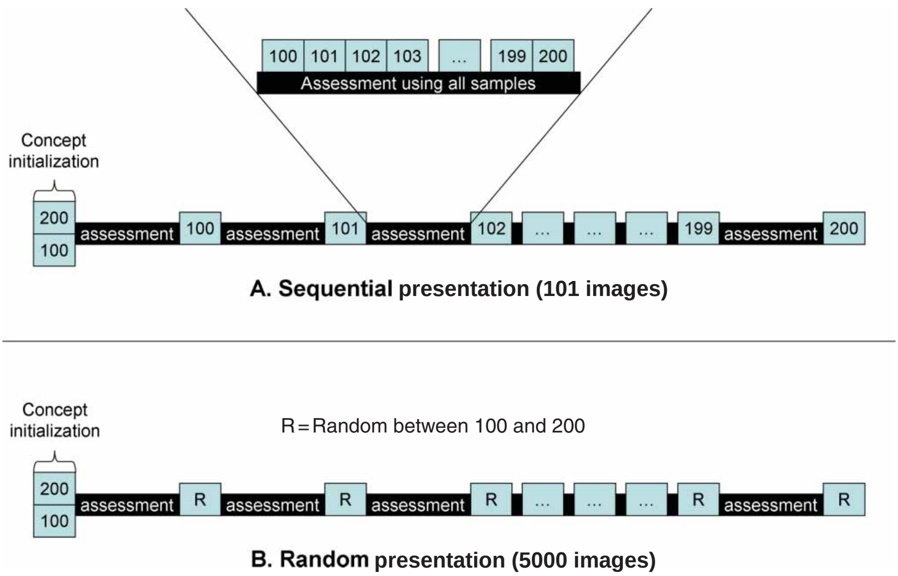

Figure 2. The presentation protocols used. The Sequential Presentation Protocol (2A) and the Random Presentation Protocol (2B) included interlaced assessment periods following each processed sample.

Table 1. The assessment process.

Sample proximity	Concept centre
$SP_{i,k} \longleftarrow \dfrac{1}{1 + \lVert s_i - w_k \rVert}$	$\mathrm{CC}_k \longleftarrow \underset{i}{\operatorname{argmin}} \lVert s_i - w_k \rVert$

Note: s_i (for any $i = 100..200$) are the morphing sequence samples. W_k (for $k \in \{`A', `B'\}$) are the learner concepts. $SP_{i,k} \in (0, 1]$ – the proximity recorded between each sample s_i and concept k. The proximity approaches 1 as the sample approaches the concept. The proximity approaches 0 as the sample Euclidean distance from the concept approaches infinity. CC_k – the recorded sample index closest to concept k

At first, CF2 was initialised with the eight shapes. Then an assimilation phase was used in which the shapes were repeatedly introduced 100 times, allowing learning and unlearning to take place. This phase was used to allow any similar shapes to effect each other. Last, a morph sequence between a square and a circle was introduced.

4. Results

The CF1 and CF2 algorithms were repeatedly evaluated while using different parameters. Each time, the learner was exposed to either a sequential presentation or a random presentation of samples from the morphing sequence. The evidence collected show that during the sequential presentation protocol, both CF1 and CF2 followed Concept 'A' as it gradually drifted towards Concept 'B'. During the random presentation protocol, both algorithms well behaved as concepts remained relatively stable and no concept overtook the complete morphing space. Similar results were found using different sets of parameters. The difference between CF1 and CF2 was demonstrated. It was shown that using CF1, a concept does not drift to a region previously learned as belonging to a different concept. It was further shown that CF2 is free from such a limitation.

4.1. *Supervised incremental learning using a sliding window approach*

As a reference, we first experimented with an incremental supervised learner that uses a simple sliding window approach. The 10 latest labels of each concept were equally weighted and used to determine the concept centre. We tested under the assumption that not all samples are labelled. Figure 3 shows the results when using 10%, 2.5% and 0% of samples being labelled. Note that substantial adaptation of the concept was exhibited with 10% of samples being labelled. Yet, Concept 'B' was not unlearned and images 179–200 remain closer to Concept 'B' than to Concept 'A'. As shown in the mid and left plots, the percentage of samples being labelled modulates the adaptation of Concept 'A'.

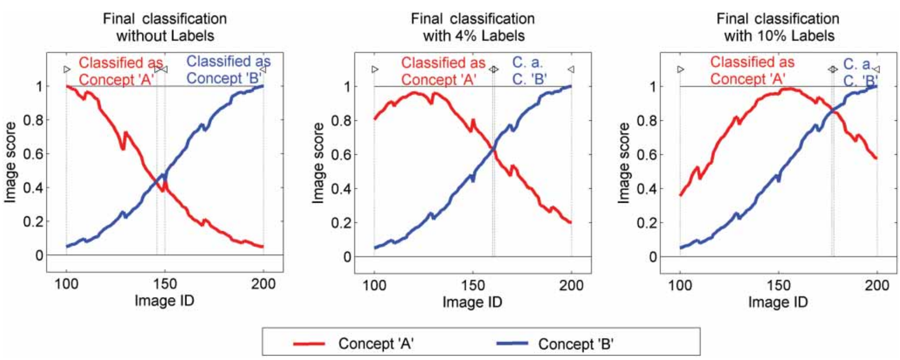

Figure 3. Final sample proximity using a supervised sliding window approach. The three plots show the results of a moving window incremental supervised learner that use an average of the 10 latest labels per concept. The right plot shows the results when 10% of the samples are labelled. Concept 'A' is closer to most samples while Concept 'B' is closer to samples 179–200. Concept 'A' shifted towards Concept 'B' and is now closest to image 156. The mid plot shows the results when 2.5% of the samples are labelled. Concept 'A' and Concept 'B' roughly divide the morphing space. Concept 'A' slightly shifted towards Concept 'B' and is now closest to image 130. The left plot shows the results where none of the samples are labelled. Concept 'A' and Concept 'B' roughly divide the morphing space. Concept 'A' had not shifted towards Concept 'B'.

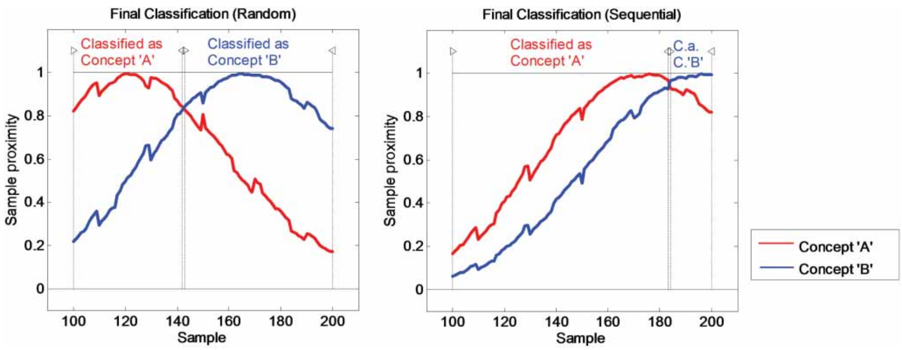

Figure 4. Final sample proximity using CF1 ($\eta = 0.1$, $\theta = 100$). The right subplot shows the proximity of samples to each concept after CF1 was presented with the complete sequential presentation protocol; Concept 'A' is closer to most samples while Concept 'B' is closer to samples 184–200. The left subplot shows the proximity of samples to each concept after CF1 was presented with the complete random presentation protocol; Concept 'A' and Concept 'B' divide the morphing space at approximately midway.

4.2. *Concept Follower algorithm (CF1)*

The CF1 algorithm was used with a variety of learning rate (η) and threshold (θ) parameters. Figure 4 shows one exemplar of the results for a case where the learning rate parameter is $\eta = 0.1$. The threshold (θ) was set to 100 (under a 551 dimensional space in which the difference between a white pixel and a black pixel is 255). After completing the presentation of all sequential protocol samples, the proximity of each morphing sequence sample to each concept was calculated. The right plot shows the assessment results. Samples at the range 100–183 are closer to Concept 'A'. Samples at the range 184–200 are closer to Concept 'B'. Following the sequential presentation protocol, Concept 'B' is closest to sample 196 not far from the starting point prior to the protocol (sample 200); Concept 'A' is closest to sample 176 and drifted significantly from sample 100 where it was initialised. Similar results were found for a large range of learning rates and thresholds. The left plot shows the assessment results following the random presentation protocol. Samples at the range 100–142 are closer to Concept 'A'. Samples at the range 143–200 are closer to Concept 'B'.

Figure 5 shows the assessment results following each step of the incremental learning. Following each assessment, the sample closest to the Concept was defined as the concept centre. The right plot shows the concept centres during the sequential presentation. It can be seen that Concept 'A' drifts towards the region occupied by Concept 'B'. Yet, as the change approaches the prelearned Concept 'B', once sample 183 is presented, the learner stops drifting Concept 'A' towards Concept 'B'. Instead Concept 'B' is drifted towards the centre and later back towards sample 200. Note that the figure shows only the sample closest to the concept as a measurement of the concept overall drift. Not all concept changes are shown. As a comparison, the left plot shows the concept centres during the random presentation. It can be seen that Concepts 'A' and 'B' quickly drifts towards the centre and then remain relatively stable during the 5000 sample presentation. Throughout the random presentation, the two concepts divide the morphing space between sample 100 and 200 to approximately two equal halves.

4.3. *Concept Follower with Unlearning algorithm (CF2)*

The CF2 with Unlearning algorithm was used with a variety of learning rate (η), unlearning rate (δ) and threshold (θ) parameters. Figure 6 shows one exemplar of the results for a case where the learning rate parameter is $\eta = 0.1$ and the unlearning rate parameter is $\delta = 0.05$. The threshold (θ) was set to 100 (under a 551 dimensional space in which the difference between a white pixel

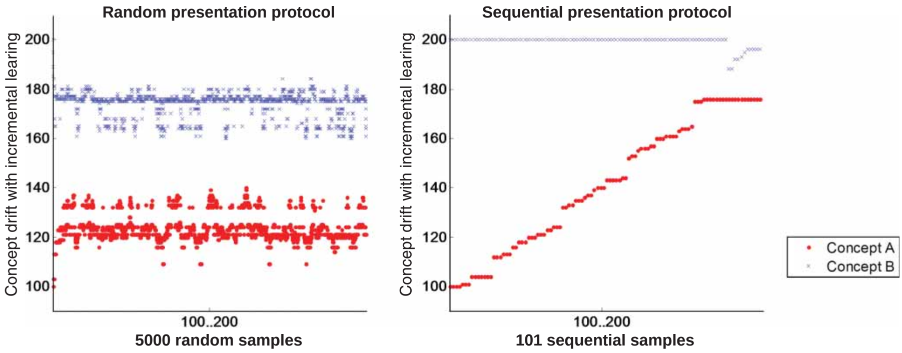

Figure 5. Concept centre drift results using CF1 ($\eta = 0.1$, $\theta = 100$). The right subplot shows the drift of Concept 'A' centre during the presentation of 101 samples representing a morphing sequence from sample 100 and sample 200; Concept 'A' gradually drifts towards Concept 'B'. Once the two Concepts overlap, Concept 'A' discontinues its drift. The left subplot shows the drift of concept centres during the presentation of 5000 samples, equally distributed along the morphing sequence between sample 100 and sample 200; Concept 'A' and Concept 'B' divide the morphing space at approximately midway.

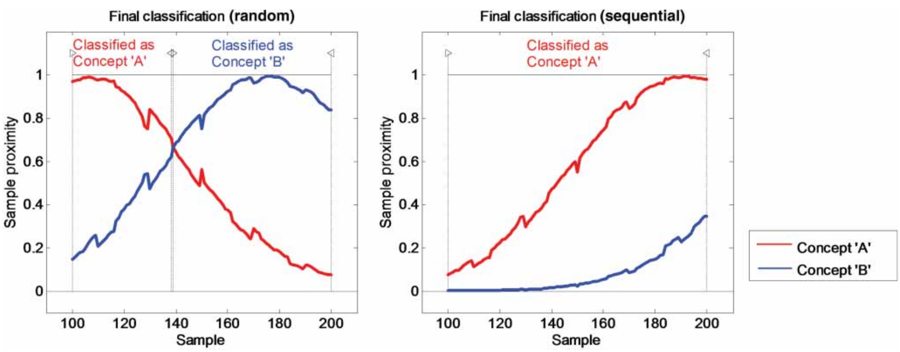

Figure 6. Final sample proximity using CF2 ($\eta = 0.1$, $\delta = 0.05$, $\theta = 100$). The right subplot shows the proximity of samples to each concept after CF2 was presented with the complete sequential presentation protocol; Concept 'A' is closer to all possible 101 samples. The left subplot shows the proximity of samples to each concept after CF2 was presented with the complete random presentation protocol; Concept 'A' and Concept 'B' equally divide the morphing space.

and a black pixel is 255). After completing the presentation of all sequential protocol samples, the proximity of each morphing sequence sample to each concept was calculated. The right plot shows the assessment results. Samples at the range 100–200 were all classified as closer to Concept 'A'. Concept 'B' was unlearned as the Concept representing sample 200. Following the sequential presentation protocol, Concept 'B' is still closest to sample 200 but not as close as Concept 'A'; Concept 'A' is closest to sample 191 and drifted significantly from sample 100 where it was initialised. Similar results were found for a large range of learning rates and thresholds. The left plot shows the assessment results following the random presentation protocol. Samples at the range 100–138 are closer to Concept 'A'. Samples at the range 139–200 are closer to Concept 'B'.

Figure 7 shows the assessment results following each step of the incremental learning. Following each assessment, the sample closest to the concept was defined as the concept centre. The right plot shows the concept centres during the sequential presentation. It can be seen that Concept 'A' drifts towards the region occupied by Concept 'B'. Yet, as the change approaches the prelearned Concept 'B', once sample 174 is presented, the learner drifts Concept 'B' away such that sample 200 is no longer closest to Concept 'B' but becomes closer to Concept 'A'. As a comparison, the left plot shows the concept centres during the random presentation. It can be seen that Concepts

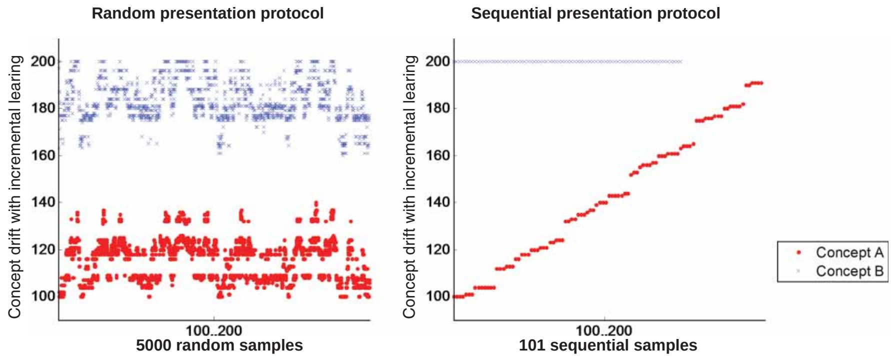

Figure 7. Concept centre drift results using CF2 ($\eta = 0.1$, $\delta = 0.05$, $\theta = 100$). The right subplot shows the drift of Concept 'A' centre during the presentation of 101 samples representing a morphing sequence from sample 100 and sample 200; Concept 'A' gradually drifts towards Concept 'B'. Once the two Concepts overlap, Concept 'A' continues its drift towards sample 200 and Concept 'B' drifts away from the morphing sequence. The left subplot shows the drift of concept centres during the presentation of 5000 samples, equally distributed along the morphing sequence between sample 100 and sample 200; Concept 'A' and Concept 'B' divide the morphing space at approximately midway.

'A' and 'B' quickly drifts towards the centre and then remain relatively stable during the 5000 sample presentation. Throughout the random presentation, the two concepts divide the morphing space between sample 100 and 200 to approximately two equal halves.

4.4. *Sensitivity of CF1 and CF2 to parameters*

The sensitivity of the presented protocols to the learning rate (η), threshold (θ), and unlearning rate (δ) parameters were evaluated. We used CF1 with $\theta = 100$ to evaluate learning rates (η) starting at $\eta = 0.001$ and factored by 1.2 per each iteration till we reached $\eta = 0.85$. We used CF1 with $\eta = 0.1$ to evaluate thresholds starting at $\theta = 1$ and factored by 1.2 per each iteration till we reached $\theta = 1021$. We used CF2 with $\eta = 0.1$ and $\theta = 100$ to evaluate unlearning rates starting at $\eta = 0.001$ and factored by 1.2 per each iteration till we reached $\eta = 0.7$.

Figures 8–10 show the results on a logarithmic scale. The left plot per each figure, shows the percentage of errors (E divided by the number of samples per presentation protocol) for a Random Presentation Protocol with 1000 samples and a Sequential Presentation Protocol with 101 samples. The middle plot shows the standard deviation of the concept centres when facing a Random Presentation Protocol (the measurements are taken along the second 500 random samples in each iteration). The right plot shows the final location of the concepts following a Sequential Presentation Protocol.

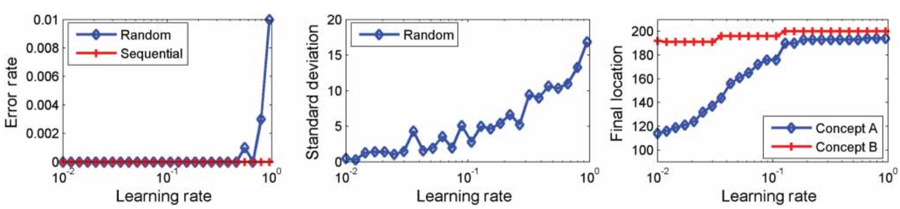

Figure 8. The effect of the learning rate parameter on CF1 ($\eta = 0.001 \ldots 0.85$, $\theta = 100$). The left subplot shows the error rate during a Random and a Sequential Presentation Protocols. The middle subplot shows the stability of the concept as evident from the standard deviation of the concept centres during a Random Presentation Protocol. The right subplot shows the final location of the two concepts following a Sequential Presentation Protocol. As shown, setting CF1 with learning rates in the range of $0.05 < \eta < 0.5$ allows us to avoid errors, offer an approximately constant stability and reach similar final setup of concept centres.

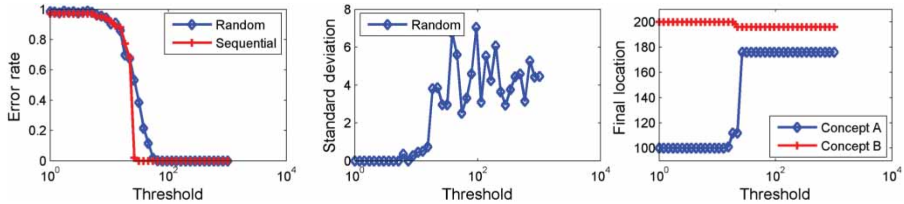

Figure 9. The effect of the threshold parameter on CF1 ($\theta = 1 \ldots 1021$, $\eta = 0.1$). The left subplot shows the error rate during a Random and a Sequential Presentation Protocols. The middle subplot shows the stability of the concept as evident from the standard deviation of the concept centres during a Random Presentation Protocol. The right subplot shows the final location of the two concepts following a Sequential Presentation Protocol. As shown, setting CF1 with thresholds in the range of $80 < \theta < 1000$ allows us to avoid errors, offer an approximately constant stability and reach the same final setup of concept centres.

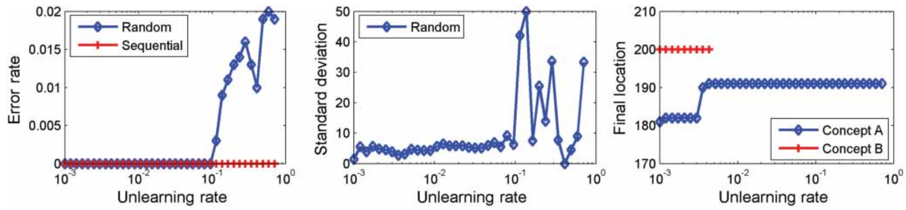

Figure 10. The effect of the unlearning rate parameter on CF2 ($\delta = 0.001 \ldots 0.7$, $\eta = 0.1$, $\theta = 100$). The left subplot shows the error rate during a Random and a Sequential Presentation Protocols. The middle subplot shows the stability of the concept as evident from the standard deviation of the concept centres during a Random Presentation Protocol. The right subplot shows the final location of the two concepts following a Sequential Presentation Protocol. As shown, setting CF2 with unlearning rates in the range of $0.005 < \delta < \eta$ allows us to avoid errors, offer an approximately constant stability and reach the same final setup of concept centres.

It is shown that in the range $0.05 < \eta < 0.5$, no errors appear, and Concept 'A' drifts more than halfway towards Concept 'B'. Note that an increase in the learning rates results in an increase in the standard deviation, suggesting greater concept instability. It is shown that for $\theta > 80$, no errors appear, and the concept behaviour is identical. Note that in this range, no specific pattern appears in the standard deviation, suggesting constant stability. It is shown that in the range $0.005 < \delta < \eta$, no errors appear, and the concept behaviour is identical. Note again the constant stability.

These parametric sensitivity results show that for a large choice of parameters, CF1 and CF2 are stable and the used parameters have little effect on the qualitative results.

4.5. *Multi-shape evaluation results*

The effect of CF2 on a group of pre-learned shape concepts was evaluated. The learning rate parameter was set to $\eta = 0.2$ while the unlearning parameter was set to $\delta = 0.1$. The threshold (θ) was set to 1000 (under a 1600 dimensional space in which the difference between a white pixel and a black pixel is 255).

Figure 11, top row shows the eight concepts after CF2 was initialised using the eight shapes. The middle row shows the eight concepts following an assimilation phase in which the shapes were repeatedly introduced. Note the circle and the half circle concepts were slightly modified such that following the assimilation phase, the distance between the two concepts increased. Following a morph phase in which CF2 was exposed to a square shape being morphed into a circle shape, we note that the square concept now have a circle shape. The circle concept was also changed. Following these changes, the circle shape is now more similar to the square concept compared to the circle concept. Indeed, the circle shape is now classified as belonging to the square concept.

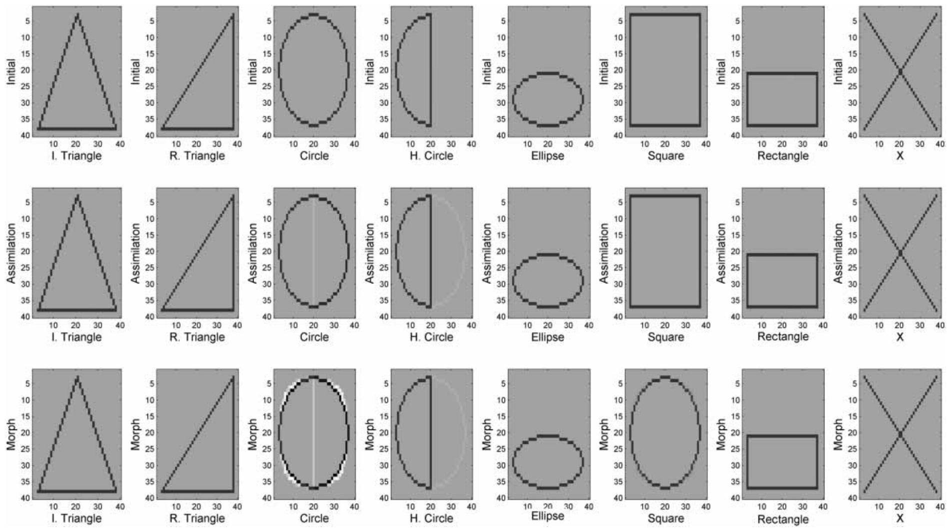

Figure 11. Shape morping ($\eta = 0.2$, $\theta = 1000$, $\delta = 0.01$). Eight concepts of CF2 were initialised using eight different shapes as shown in the top row. Then, an assimilation phase was used in which the eight shapes were repeatedly introduced. The results of the assimilation phase are shown in the middle row. Note the effect of unlearning on the two similar H. circle and circle concepts. Last, a morph between a square and a circle was introduced and the results are shown in the bottom row. Note that following the learning that occurred during the exposure to the morph sequence, the square concept is now changed to a circle shape. Note also that following unlearning, the original circle concept was also changed.

5. Discussion

This paper was motivated by a biologically observed strategy in which unlabelled samples were successfully used to incrementally adapt pre-learned concepts, while coping with dramatic environmental changes. Two unsupervised incremental *CF* algorithms (CF1 and CF2) were developed to follow prelearned concepts.

The presented CFs were adapted from the *LF* algorithm. Using CF1 and CF2, a machine learner can tune concepts when facing gradual environmental changes. While deriving CF1 and CF2 from LF, we have intentionally removed the ability of LF to learn new concepts. Under the strategy devised here, Concept Followers are used only for the purpose of adjusting pre-learned concepts. This behaviour is in line with the requirement for concept structure stability. Adhering to such structural stability, the structure learned by the supervised learner should be preserved by the Concept Follower. Therefore, we do allow the CFs to learn new concepts. Also, a supervised learner may use sample labels to derive a meaningful label to learned concepts. Concepts Followers do not use labelled samples and may not derive such meaningful label to learned concepts. Instead of learning new concepts, we designed CF1 and CF2 to detect a newly formed concept as an abrupt change of the environment, requiring re-initiation of the supervised phase.

The experimental layout included a controlled concept drift, an artificially created morphing sequence between two faces (Wallis and Bulthoff 2001; Preminger et al. 2007; Hadas et al. 2010). The sample presentation was unbalanced; one gradually changing pre-learned concept was displayed, while a second pre-learned concept was hidden. Experiments showed that CF1 and CF2 can drift the changing concept inline with changes in the stimuli even when the changes accumulate. During unbalanced presentation, CF1 was unable to drift the changing concept to a region previously occupied by the hidden concept; CF2 was able to overcome such limitation and followed a pattern more similar to the one observed in biological systems.

Since the developed CF1 and CF2 are based on the LF algorithm, they share its limitations. CF1 and CF2 can be used by a learner when assuming the same sample probability density function (PDF) to all concepts; and when the PDF is assumed to be monotonic and spherical, i.e., the PDF can be approximated by observing the Euclidean distance between the sample and pre-learned concept centres. It is further assumed that initial concept centres can be determined by a supervised learning stage. Such limitations of CF1 and CF2 may be lifted by developing CFs based on other unsupervised incremental learning algorithms.

Unsupervised CFs are useful in problem areas in which the concept behaviour over time can be inferred from the sample distribution. The accuracy of many unsupervised algorithms including LF is increased when: (1) Concepts are fairly sparse in the sample space; (2) Concepts are characterised by having a sharp PDF boundary area where the probability of samples inside the area is significantly higher than the sample probability outside the area. Such conditions are more likely to exist in a high-dimensional sample space.

Physiological and psychophysical studies demonstrated how biological systems may learn concepts using labelled samples and later adapt such concepts using unlabelled samples (Wallis and Bulthoff 2001; Leutgeb et al. 2005,Preminger et al. 2007, Hadas et al. 2010). Classifying unlabelled stimulus affects future classifications of resembling stimuli. The system adapts incrementaly and follow the environmental changes. Using this mechanism, the system can radically drift pre-learned concepts and accumulate environmental changes. CF1 and CF2 algorithms exhibit similar performance in machine learners and drift the concepts following environmental changes. CF1 and CF2 drift concepts using unlabelled samples and affect future classifications of stimuli resembling the classified stimulus. If the environment change continues, then CF1 and CF2 can accumulate change and continue to follow the environment.

Biological systems are also able to drift a first concept such that it would completely overlap a region previously learned as belonging to a second concept. This was demonstrated with unbalanced sample distribution between concepts where samples from the second concept were not presented (Hadas et al. 2010). We tested CF1 and CF2 under similar conditions. Although CF1 is able to drift the first concept and accumulate change towards the second concept, it was shown that when the first concept reaches the region in which the second concept resides, additional accumulation of drift is not enabled. Unlike CF1, CF2 continues to accumulate drift when exposed to the same pattern. Using CF2, a pre-learned concept could accumulate changes such that it can replace a second pre-learned concept, inline with the results gathered from biological systems.

Note that some problem areas would find unlearning, as used in CF2, undesirable. Under an unbalanced sample ratio a hidden concept may be unlearned even without concept drift. This may occur when one concept is very close to a second, hidden concept. In such cases CF1 which does not use unlearning may be a better candidate than CF2. As shown in the experimental results, under unbalanced sample ratio very close concepts would not merge or overlap under CF1. Dense problem areas, in which sample density of boundary areas between overlapping concepts is high, may find CF2 not useful. CF2 may be more useful where sample ratio between concepts is balanced and/or in problem areas where concepts are expected to be well distinguished and separated by low density boundary areas.

Here, we introduced an incremental algorithm to follow changing pre-learned concepts without labelled samples. The presented unsupervised methods aid a learner cope with gradual concept drift by modifying the learner concepts. The use of the CF1 and CF2 algorithms is suggested primarily for problem areas in which labels are not available in a timely manner. Supervised incremental learning methods rely on a constant stream of labels and are therefore unsuitable to adjust a learner to concept drift in the said problem areas. This is demonstrated here using the reference sliding window supervised incremental learner.

We argue that even in problem areas where labels are available but are scarce, unsupervised methods for concept drift offer advantages over supervised ones. Such advantages include: (1)

Using supervised methods, significant time may elapse till labelled samples become available and until tuning the machine learner can take place. During such time the machine classification accuracy is reduced. Where changes are frequent, supervised methods leads to a constant accuracy penalty. Unsupervised CFs help tune a machine learner shortly after a change, leading to a smaller accuracy penalty after each change. Depending on the availability of labelled samples, unsupervised CFs may offer more accurate classification over supervised methods. (2) Certain gradually changing patterns may appear to a supervised learner as an abrupt change. In some problem areas, labelled samples are only available for some of the samples but not for all. When changes occur more frequently than the rate in which new labelled samples become available, accumulated changes may go undetected and when the next labelled sample appears it may represent an abrupt change. Depending on the supervised method used, abrupt changes may not allow incremental learning (Kuncheva 2004). The frequency of labelled samples and of changes may therefore, limit the suitability of certain incremental supervised learning methods. Unsupervised CFs do not suffer from a similar limitation since all samples are used for learning. Learning can therefore take place before gradual changes accumulate into an abrupt change. (3) Where labelled samples are scarce, supervised learners face a greater challenge of identifying true environmental changes from noise. Unsupervised CFs utilise all samples for learning and therefore may offer better robustness to noise under certain problem areas. Additional research is needed to evaluate the trade off between label frequencies and its effect of the choice between supervised and unsupervised incremental learning methods for coping with concept drift.

Unsupervised CFs adjust machine learner concepts based on the distribution of the samples. Under certain problem areas, this may serve as a disadvantage compared to supervised methods. One may use labelled samples to evaluate the correctness of unsupervised Concept Followers. Alternatively, labelled samples can be used in combination with unlabelled samples to drift the learner concepts (See semi-supervised learning, for example, Zhu 2005). Another disadvantage of the presented unsupervised CFs, compared to supervised incremental learning methods is that unsupervised methods rely on each juncture to be well correlated to a previous juncture, meaning that it is suitable for coping with gradual change but not with abrupt change.

Here, we evaluated the ability of the CFs to adjust to gradual change and qualitatively compared it to the performance exhibited by the human visual system. We therefore used a data set without abrupt changes and evaluated the ability of a single unsupervised learner to cope with a gradual change. Future work may explore the use of unsupervised CFs in combination with specific supervised learners and in the context of abrupt changes.

While CF1 and CF2 are both stable in the sense of Stability-Plasticity Dilemma (Carpenter and Grossberg 1988), CF2 further adjusts other nearby concepts away from the sample; this decreases the likelihood that similar samples will be classified as one of these nearby concepts and further contributes to the stability of the classifier in the region neighboring the sample. Although stable, CF1 and CF2 exhibit significant elasticity; during the sequential presentation pattern, CF1 and CF2 have shown significant elasticity as the displayed concept accumulated changes. CF2 have shown additional elasticity as the hidden concept was changed to allow drifting the displayed concept.

The stability and elasticity tradeoff of CF1 and CF2 were controlled using the learning rate parameter (for both CF1 and CF2) and the unlearning rate parameter (for CF2). It is shown that CF1 and CF2 behave well, and offer qualitatively similar results, for a large range of parameter choices. The learning and unlearning rate parameters can be used to speed or slow the CFs. CF1 and CF2 use a threshold parameter to determine the proximity of samples affecting the pre-learned concepts. It is shown that under ideal conditions, in which abrupt change do not occur, increasing the threshold parameter do not affect the results. However, such an increase reduces the ability of CF1 and CF2 to detect abrupt changes in the environment and may result in unnecessary unlearning in CF2. In some problem areas, the threshold parameter can be determined during the initial supervised stage, by observing the sample distribution.

The multi-shape evaluation also demonstrates the stability and elasticity characteristics of the CF algorithms. It is shown that the learning and unlearning have a local effect and can therefore be used in a multi-concept environment.

In this paper, we present a first introduction into Concept Followers and characterise a novel computational model based on psychophysical findings. We evaluate the extent in which the behaviour of the computational model is similar to the one displayed by the biological model. The data set chosen was derived from the one used by the psychophysical experiments. The results show that the machine learner behaves qualitatively the same as the biological one. The use of additional data sets may help clarify the problem areas, where the suggested computational model is useful. Yet, such insight would require further analysis into the limitations of incremental supervised learning, and the introduction of performance criteria for comparing the different options. We leave this to future work.

6. Conclusion

We presented CF1 and CF2, unsupervised incremental learning algorithms based on the LF algorithm, for adapting concepts in a changing environment. It was argued that depending on the problem area, the use of unsupervised CFs such as CF1 and/or CF2 may offer significant advantages compared to alternative incremental supervised learning methods. The experimental results suggest that CF1 and CF2 are both stable and elastic. Yet such characteristics depend on the problem area sample distribution. CF2 which adds unlearning to the traditional Leader Follower mechanism was shown to follow concept drift without limitations even when the sample distribution is unbalanced. The CF2 results are inline with the performance exhibited by biological systems.

References

Carbonell, J., Fink, E., Jin, C., Gazen, C., Mathew, J., Saxena, A., Satish, V., Ananthraman, S., Dietrich, D., Mani, G., Tittle, J., and Durbin, P. (2006), 'Scalable Data Exploration and Novelty Detection', in *a NIMD Workshop*.

Carpenter, G.A., and Grossberg, S. (1988), 'The ART of Adaptive Pattern Recognition by a Self-Organizing Neural Network', *IEEE Computer*, 21(3), 77–88.

Chen, S., and He, H. (2009), 'SERA: Selectively Recursive Approach Towards Nonstationary Imbalanced Stream Data Mining', in *Neural Networks, 2009. IJCNN 2009. International Joint Conference on*, pp. 522–529.

Ditzler, G., Muhlbaier, M., and Polikar, R. (2010), 'Incremental Learning of New Classes in Unbalanced Datasets: Learn++.UDNC', Vol. 5997, eds. N. El Gayar, J. Kittler and F. Roli, Heidelberg: Springer Berlin, pp. 33–42.

Duda, R.O., Hart, P.E., and Stork, D.G. (2000), *Pattern Classification*, New York: Wiley-Interscience Publication.

Elwell, R., and Polikar, R. (2009), 'Incremental Learning of Variable Rate Concept Drift', in *Proceedings of the 8th International Workshop on Multiple Classifier Systems*, Reykjavik, Iceland, MCS '09, Berlin, Heidelberg: Springer, pp. 142–151.

Fan, W., Wang, T., Bouguet, J.Y., Hu, W., Zhang, Y., and Yeung, D.Y. (2007), 'Semi-supervised Cast Indexing for Feature-Length Films', in *Proceedings of the 13th International MultiMedia Modelling Conference*, MMM-2007, Singapore, pp. 625–635.

Gao, J., Fan, W., and Han, J. (2007), 'On Appropriate Assumptions to Mine Data Streams: Analysis and Practice', in *Proceedings of the 2007 Seventh IEEE International Conference on Data Mining*, Washington, DC, USA: IEEE Computer Society, pp. 143–152.

Garg, M., and Shyamasundar, R.K. (2004), 'A Distributed Clustering Framework in Mobile Ad Hoc Networks', in *International Conference on Wireless Networks*, pp. 32–38.

Gennari, J.H., Langley, P., and Fisher, D. (1989), 'Models of Incremental Concept Formation', *Artificial intelligence*, 40(1–3), 11–61.

Gersho, A., and Yano, M. (1985), 'Adaptive Vector Quantization by Progressive Codevector Replacement', in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, ICASSP-1985, Tampa, Florida, USA, pp. 133–136.

Hadas, D., Intrator, N., and Yovel, G. (2010), 'Rapid Object Category Adaption During Unlabelled Classification', *Perception*, 39(9), 1230–1239.
Helmbold, D.P., and Long, P.M. (1994), 'Tracking Drifting Concepts by Minimizing Disagreements', *Machine Learning*, 14(1), 27–45.
Hulten, G., Spencer, L., and Domingos, P. (2001), 'Mining Time-Changing Data Streams', in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD-2001, San Francisco, California, USA, pp. 97–106.
Kelly, M.G., Hand, D.J., and Adams, N.M. (1999), 'The Impact of Changing Populations on Classifier Performance', in *KDD '99: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA*, NY, USA: ACM Press, pp. 367–371.
Klinkenberg, R. (2004), 'Learning Drifting Concepts: Example Selection vs. Example Weighting', *Intelligent Data Analysis*, 8(3), 281–300.
Klinkenberg, R., and Joachims, T. (2000), 'Detecting Concept Drift with Support Vector Machines', in *Proceedings of the Seventeenth International Conference on Machine Learning (ICML)*, Los Altos, CA: Morgan Kaufmann, pp. 487–494.
Klinkenberg, R., and Renz, I. (1998), 'Adaptive Information Filtering: Learning in the Presence of Concept Drifts', in *Workshop Notes of the ICML/AAAI-98 Workshop Learning for Text Categorization*, Silver Spring, MD: AAAI Press, pp. 33–40.
Kong, S.G., and Kosko, B. (1991), 'Differential Competitive Learning for Centroid Estimation and Phoneme Recognition', *IEEE Transactions on Neural Networks*, 2(1), 118–124.
Kosko, B. (1991), 'Stochastic Competitive Learning', *IEEE Transactions on Neural Networks*, 2(5), 522–529.
Kuncheva, L.I. (2004), 'Classifier Ensembles for Changing Environments', in *Proceedings of the 5th International Workshop on Multiple Classifier Systems*, Springer, Berlin, pp. 1–15.
Kuncheva, L.I. (2008), 'Classifier Ensembles for Detecting Concept Change in Streaming Data: Overview and Perspectives', in *2nd Workshop SUEMA 2008*, Greece, pp. 5–10.
Leutgeb, J.K., Leutgeb, S., Treves, A., Meyer, R., Barnes, C.A., McNaughton, B.L., Moser, M.B., and Moser, E.I. (2005), 'Progressive Transformation of Hippocampal Neuronal Representations in m̈orphedË nvironments', *Neuron*, 48, 345–358.
Liu, D., and Kubala, F. (2004), 'Online Speaker Clustering', in *Proceedings of the 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, ICASSP-2004, Quebec, Canada, Vol. 1, pp. 333–336.
Maloof, M.A., and Michalski, R.S. (2004), 'Incremental Rule Learning with Partial Instance Memory', *Artificial Intelligence*, 154, 95–126.
Natale, C.D., Fabrizio, A.M.D., and D'Amico, A. (1995), 'A Self-Organizing System for Pattern Classification: Time Varying Statistics and Sensor Drift Effects', *Sensors and Actuators B: Chemical*, 27(1–3), 237–241, Eurosensors VIII.
Nunez, M., Fidalgo, R., and Morales, R. (2007), 'Learning in Environments with Unknown Dynamics: Towards more Robust Concept Learners', *Journal of Machine Learning Research*, 8, 2595–2628.
Pei, X.m., Xu, J., Zheng, C., and Bin, G.y. (2005), 'An Improved Adaptive RBF Network for Classification of Left and Right Hand Motor Imagery Tasks', *Lecture Notes in Computer Science*, pp. 1031–1034.
Preminger, S., Sagi, D., and Tsodyks, M. (2007), 'The Effects of Perceptual History on Memory of Visual Objects', *Vision Research*, 47(7), 965–973.
Salganicoff, M. (1997), 'Tolerating Concept and Sampling Shift in Lazy Learning Using Prediction Error Context Switching', *AI Review, Special Issue on Lazy Learning*, 11(1–5), 133–155.
Tsymbal, A. (2004), 'The Problem of Concept Drift: Definitions and Related Work', Technical report, Trinity College Dublin, Ireland.
Tsypkin, Y.Z. (1973), *Foundations of the Theory of Learning Systems*, New York: Academic Press.
Wallis, G., and Bulthoff, H.H. (2001), 'Effects of Temporal Association on Recognition Memory', *Proceedings of the National Academy of Sciences, USA*, 98(8), 4800–4804.
Widmer, G., and Kubat, M. (1996), 'Learning in the Presence of Concept Drift and Hidden Contexts', *Machine Learning*, 23(1), 69–101.
Zehraoui, F., and Bennani, Y. (2004), 'M-SOM-ART:Growing Self Organizing Map for Sequences Clustering and Classification', in *Proceedings of the 16th Eureopean Conference on Artificial Intelligence, ECAI'2004*, Valencia, Spain, pp. 564–570.
Zhu, X. (2005), 'Semi-Supervised Learning Literature Survey', Computer Sciences Technical Report TR 1530, University of Wisconsin-Madison.
Zliobaite, I. (2009), 'Learning under Concept Drift: an Overview', Technical report, Faculty of Mathematics and Informatics, Vilnius University: Vilnius, Lithuania.