Unsupervised Learner for coping with gradual Concept Drift

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

Computer Science

by

David Hadas

Thesis Supervisor:
Professor Nathan Intrator

April 2011
Abstract

Current computational approach to incremental learning requires a constant stream of labeled data to cope with environmental changes known as concept drift. In our research we consider a case where labeled data is unavailable.

We begin this investigation with observing how the human visual system copes with environmental change. We investigate unsupervised adaptation processes of the human visual system by using a presentation protocol that includes an ordered sequence of morphed stimuli. Recent reports from electrophysiological and psychophysical experiments provide evidence that using such a protocol, the visual system is capable of adapting its concepts without labels. A new psychophysical protocol that uses an unbalanced sample ratio is presented. We show that under the new protocol, subjects may replace a pre-learned face with an entirely different face. The results suggest that the visual system may adapt to concept drift without labels quickly and robustly.

Inspired by the performance of the human visual system, capable of adjusting its concepts using unlabeled stimuli, we devise a new strategy for machine learners to cope with gradual concept drift. Under the new strategy, concepts are learned first using a supervised learner and later modified
by an unsupervised learner, which we name a Concept Follower. We introduce two variants of Concept Followers that can adjust pre-learned concepts to environmental changes using unlabeled data samples. The Concept Followers are derived from an Unsupervised Competitive Learning algorithm known as the Leader Follower. We motivate the needed change in the existing Leader Follower algorithm and evaluate the performance of the two variants. We show that unsupervised machine learning can be used to cope with the accumulation of environmental changes while facing an unbalanced sample ratio.

David Hadas david.hadas@gmail.com
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>iii</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>v</td>
</tr>
<tr>
<td>List of Tables</td>
<td>ix</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xi</td>
</tr>
<tr>
<td>List of Algorithms</td>
<td>xiii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>xv</td>
</tr>
<tr>
<td>Dedication</td>
<td>xvii</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Background</td>
<td>1</td>
</tr>
<tr>
<td>1.2 State of the art</td>
<td>3</td>
</tr>
<tr>
<td>1.3 Evidence from psychophysical data</td>
<td>6</td>
</tr>
<tr>
<td>1.4 The approach taken</td>
<td>11</td>
</tr>
<tr>
<td>1.5 Structure</td>
<td>12</td>
</tr>
</tbody>
</table>
# Table of Contents

## 2 Psychophysical Experiments

2.1 Overview .................................. 15

2.2 Paced Morphing Experimental Setting .......................... 17

2.2.1 Stimuli .................................. 17

2.2.2 Apparatus ................................ 18

2.2.3 Subjects .................................. 19

2.2.4 Procedure ................................ 19

2.3 Experiments and Results .................................. 22

2.4 Discussion of the results .................................. 29

## 3 Computational Model Characterization Based On The Psychophysical Findings .................................. 33

## 4 Methodology .................................. 37

4.1 The traditional Leader Follower algorithm (LF) ............... 38

4.2 Concept Follower algorithm (CF1) .............................. 40

4.3 Concept Follower with Unlearning algorithm (CF2) ........... 45

4.4 Mathematical Properties .................................. 47

## 5 Computational Experiments .................................. 53

5.1 Experimental Setup .................................. 53

5.2 Results .................................. 57

5.2.1 Supervised incremental learning using a sliding window approach .................................. 58
List of Tables

5.1 The assessment process. . . . . . . . . . . . . . . . . . . . . . 56
# List of Figures

2.1 The Psychophysical Experiment Morphing Sequence . . . . . . 18  
2.2 The Psychophysical Experiment Layout . . . . . . . . . . . . . 25  
2.3 The Psychophysical Experiment Target Classification Ratio . . 27  

4.1 Expected adaptation of a learned concept . . . . . . . . . . . 48  

5.1 The morphing sequence . . . . . . . . . . . . . . . . . . . . . 54  
5.2 The presentation protocols used . . . . . . . . . . . . . . . . 56  
5.3 Two shapes morphing sequence . . . . . . . . . . . . . . . . 57  
5.4 Final Sample Proximity supervised learner . . . . . . . . . . . 59  
5.5 Final Sample Proximity using CF1 . . . . . . . . . . . . . . . 60  
5.6 Concept center drift results using CF1 . . . . . . . . . . . . . 61  
5.7 Final Sample Proximity using CF2 . . . . . . . . . . . . . . . 63  
5.8 Concept center drift results using CF2 . . . . . . . . . . . . . 65  
5.9 The effect of the learning rate parameter on CF1 . . . . . . . 66  
5.10 The effect of the threshold parameter on CF1 . . . . . . . . . 67  
5.11 The effect of the unlearning rate parameter on CF2 . . . . . . 69  
5.12 Shape morping . . . . . . . . . . . . . . . . . . . . . . . . . . 71
# List of Algorithms

1. The traditional Leader Follower algorithm: \texttt{LF()} . . . . . . . . 39
2. The Concept Follower algorithm: \texttt{CF1()} . . . . . . . . . . 41
3. Concept Follower with Unlearning algorithm: \texttt{CF2()} . . . . . 45
List of Algorithms
Acknowledgments

I would like to thank my supervisor, Prof. Nathan Interator, for guiding me through the fascinating, yet sometimes painful process of scientific research, crystallizing, formalizing and stepping ideas, and honoring the written text.

I further would like to thank, Dr. Galit Yovel, who had overseen all psychophysical aspects of this work, for guiding me through the maze of psychophysical research while seeking solid ground in the uncharted terrains of the human brain.
Acknowledgments
Dedication

This thesis is dedicated to my brother, Ran Hadas, whose untimely departure one sunny Saturday morning left us all with an unwritten will for respecting our living days.

This thesis is further dedicated to Sigal who had put up with the endless hours and our three kids

King Agur, The Brave,

King Tor, The Skilled, and

Queen Shahaf, The Wise.
Chapter 1

Introduction

1.1 Background

Gradual changes in the environment may reduce machine classification accuracy [12]. Initially, the machine may learn a set of concepts. Using such learning, when the machine is given an unlabeled sample from the environment, it may be able to predict if the sample belongs to the concepts learned. A useful learning machine would be able to classify new samples, taken from an environment, and correlate the samples to the pre-learned concepts. The machine accuracy is therefore determined by the ability to predict correct labels to new samples.

Gradual change in the environment may slowly reduce the machine accuracy. Samples once belonging to one concept may later belong to another concept. The accumulation of such changes would turn a stationary machine learner less accurate and eventually useless. This thesis 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.
Chapter 1. Introduction

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 [20]. Under certain problem areas, true labels are either hard to achieve or not available in a timely manner; see for example Kelly et al. [23] describing a bank loan application where true labels become available 24 months after a loan was granted. Under such circumstances, Kuncheva [30] suggested to use unlabeled samples to detect concept drift. Here we discuss the use of unlabeled 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; labeling the new environment and training a new learner is one approach for coping with an abrupt change of the environment [29]. 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 unlabeled samples and the correlation between junctures to tune pre-learned concepts and follow gradual environmental changes.
1.2 State of the art

Learning machines often face the problem of organizing an endless stream of data into well perceived concepts and adapting such concepts according to environmental changes. The explosion of readily available digital information, as evident in recent years, brings about a growing interest in the discovery of algorithms that will incrementally attain concepts from the information sampled. Incremental learning, also named online learning is often used when the number of samples exceeds the memory available to the classifier [16, 31]. A challenging data mining problem emerges in cases where the underlying concepts may drift. Concept drift studied by Helmbold and Long [20] is now a growing research field related to incremental learning of real world concepts that may change with time [22, 41].

In machine learning literature, learning under a concept drift is usually considered in the context of supervised learning; see, for example, performance comparison by Klinkenberg and Renz [26], method classification by Maloof and Michalski [34], and a review by Zliobaite [49]. Unsupervised learning methods are sometimes suggested for detecting concept drift (for example Kuncheva [30]) and is also mentioned in the context of sensor drift (for example Natale et al. [35]).

On-line supervised learners may use a stream of new labeled samples to gradually change with time and adapt to concept drift [22, 41]. Some supervised learners use instance selection methods in which a sliding win-
Chapter 1. Introduction

dow determines the instances used by the model; see, for example, FLORA by Widmer and Kubat [44], Time-Windowed Forgetting by Salganicoff [40], OnlineTree2 by Nunez et al. [37] and a method based on Support Vector Machines by Klinkenberg and Joachims [25]. 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 [24]. 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 [10, 29]. Some supervised learners are suitable for classifying imbalanced class distributions [5, 6]. All such supervised on-line learners depend on timely availability of labeled samples.

In some application areas, using a supervisor to label samples is costly or hard to achieve. In other application areas, labels become available long after samples are classified and possibly too late to adjust the learner to a change in the environment; See, for example, Kuncheva [29]. In such areas, a strategy is which a supervisor is used only during an initial learning phase is therefore advantageous over one requiring a constant stream of labels to support concept drift. Such an advantage can be obtained by using unsupervised learning to adapt the pre-learned concepts such that they may drift to accommodate environmental change. While current research into concept-drift focuses on supervised algorithms, in this thesis, we explore the use of unsupervised algorithms for adapting a machine learner to concept drift.

Reviewing unsupervised learning literature, one may divide the available
1.2. State of the art

algorithms by their concept stability; some unsupervised learning algorithms use a global criteria optimization 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 [7]. As an example, adding a single sample to the K-Means algorithm may change all resulting clusters; a behavior which may be deemed unacceptable when tuning concepts previously learned from labeled samples. The requirement for concept structure stability is known as the Stability-Plasticity Dilemma [4].

Incremental unsupervised learning algorithms avoid using a global criteria optimization 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. [7] but also Unsupervised Competitive Learning by Kong and Kosko [27], Kosko [28] and work by Tsypkin [42]. As new unlabeled samples arrive, unsupervised algorithms such as CLAS-SIT [14], ART [4], M-SOM-ART [47] and the Leader Follower algorithm [7], 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 [15] describing a progressive codevector replacement technique.

The Leader Follower algorithm (LF) 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 unlabeled data. Some examples include: Offline
Chapter 1. Introduction

Speaker Clustering, where LF has shown promising results compared with existing offline speaker clustering, while running much more efficiently \cite{33}; project ARGUS, where LF is part of a novelty detection mechanism \cite{3}; a distributed version of LF was suggested to enable the formation of Mobile Ad Hoc Networks \cite{13}; in Pei et al. \cite{38}, an LF clustering method is used for continuous classification of hand motor imagery tasks; Fan et al. \cite{11} 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. Instead, machine learning literature considers supervised learning methods for coping with concept drift. Such supervised learning methods, which rely on a constant stream of labeled samples, were shown to follow gradual and even abrupt concept drift. In problem areas where labels are scarce or not available in a timely manner, supervised methods may become less accurate and may not allow adaptation to concept drift. We research here the use of incremental unsupervised learning methods for coping with gradual concept drift.

1.3 Evidence from psychophysical data

The human visual system can cope well with concept drift. Biological systems use incremental learning to gradually grasp concepts hidden in an endless stimuli stream; See, for example, Wilson \cite{46} and a review by Dudai \cite{8}.
1.3. Evidence from psychophysical data

Recent electrophysiological and psychophysical studies suggest that biological systems incrementally adapt pre-learned concepts through an unsupervised learning process [32, 39, 43]. The evidence suggests that after an initial phase in which concepts were learned, the system may radically adapt the learned concepts following stimuli exposure. The biological system appears to slowly accumulate changes as learned from the stimuli such that the learned concept drifts are aligned with object changes.

Categorical perception is said to exist where (1) stimuli from a continuum are labeled as belonging to different classes and (2) stimuli labeled differently are well discriminated whereas others, labeled the same, are less well discriminated [9]. Categorical perception therefore requires that there be a sharp “perceptual boundary” between stimuli from two different categories. Monitoring changes in the perceptual boundary between two categories sheds light on the internal representation of the category [18]. A perceptual boundary appears between general categories (e.g. faces, chairs, cars, etc.), but also between specific exemplars within a general category (e.g. face stimuli from Jerry, George and Elaine). Thus all stimuli sharing the same label (e.g. Jerry) are perceived as the same object category [1, 36].

Recent results from electrophysiological and psychophysical experiments provide evidence that exposure to an ordered sequence of morphed stimuli between two pre-learned object categories adapts the perceptual boundary between the categories and suggest that exposure to a morphing sequence may expand an object category [2, 32, 39]. Leutgeb et al. [32] reported that
Chapter 1. Introduction

hippocampal place cells are capable of incremental plastic deformation. The researchers trained rats inside a square or circular enclosure in a random sequence. The rats’ task was to search for food scattered around the enclosure. Pretraining continued for 16-19 days, until the two enclosure conditions reliably activated different subsets of neurons in the associative CA3 network. Cell ensemble activity was then recorded while the square box was morphed into a circle, and vice versa, through a sequence of five intermediate shapes. The researchers showed that during exposure to a gradually morphed enclosure, the cell ensemble firing patterns of intermediate shapes depended on the direction of the morphing (stimuli order). The researchers then repeated the sequence in alternating directions on subsequent days and reported that the difference between the firing patterns of the previously established end points had attenuated. In another hippocampal place cells recording experiment, rats were pretrained on similar stimuli but later exposed to a scrambled order of intermediate shapes. While exposure to the gradually morphed stimuli resulted in attenuated differences in firing patterns, exposure to the scrambled order, instead, resulted in an abrupt switch between firing patterns of square-like and circle-like intermediate shapes [45].

The hippocampal place cells evidence suggests that exposure to a concept (e.g. square) that gradually drift to another (circle) result in attenuation of the firing patterns distinguishing between the two pre-learned concepts. Exposure to a random sequence do not result in such change. The observed stimuli order effect is discussed in Blumenfeld et al. [2], who showed that
1.3. Evidence from psychophysical data

attractor networks [21] may also display a similar effect. The researchers modulated the Hebbian learning mechanism [19] of the attractor network using weights determined by the perceived "novelty" of the stimuli. Using such novelty-facilitated modulation, the researchers showed that when morph patterns are presented in a gradually increasing order, the network exhibits a stronger tendency for merged representations compared to random order. The model, therefore, predicts that when network inputs are correlated, as in the case of a morphing sequence, different prelearned representations may expand and collapse into a single unified representation. A similar stimuli order effect was also observed during psychophysical experiments. Wallis and Bulthoff [43] conducted an experiment in which subjects were exposed to rotated heads, artificially constructed through morphing from a face in a frontal view and another in a profile view. Using discrimination tests, the researchers showed that the subjects associated the two faces as belonging to the same object. Subjects therefore associated different faces as belonging to the same object following a concept drift from one face to the other. When subjects were exposed to the morphed head images in a randomized order, no association was discovered. The authors suggested that spatiotemporal correlation was required to associate the face representations.

In a recent study, Preminger et al. [39] have shown that associations can be obtained using spatial correlation, even when temporal correlation is avoided. In their study, subjects initially learned to classify a group of facial images as “friends” and others as “nonfriends.” A sequence of morphed
images was generated between a “friend” and a “nonfriend.” At the onset, the subjects associated about half of the morphed sequence as a “friend.” In other words, the object category boundary between a friend and a nonfriend was located around the 50-50% morphed image. In 10 of 18 cases, exposure to the ordered sequence repeatedly, over many days, resulted in a gradual change in which more of the intermediate morphed images were classified as a “friend.” Finally, after repetitive sessions, these subjects identified the two morphed faces as “friends.” Exposure without labels to the morph sequence led to a significant drift in one pre-learned concept such that it was changed to include a space once occupied by a different pre-learned concept. The authors called this change in the object boundary a “morph effect.” In the remaining 8 of the 18 subjects a morph effect was not detected. The researchers indicated that the subjects who did not experience the morph effect had initially been more discriminative in identifying the morphed images as a “nonfriend.” A control group exposed to a random unordered presentation of the same morphed images did not produce a morph effect. Note that when preparing the morphing sequence, the researchers ensured that there were no conspicuous features to cause trivial discrimination between faces. The constructed faces were chosen to be neither too similar nor too different from each other.

Taken together, the above reports suggest that a biological system exposed to a Concept Drift from a first concept to a second, adjusts to the change. Such an adjustment results in the second concept being classified the
1.4 The approach taken

same as the first concept. The evidence further suggests that when a biological system classifies a stimulus as belonging to a concept it may also undergo a process of unsupervised or internally supervised learning that affects future classifications of stimuli resembling the classified stimulus. This incremental learning process, allows the system to gradually drift its concepts without a supervisor such that they can track significant environmental changes. In particular, biological systems are able to handle concept drift even if the drift is toward a space that was previously learned as belonging to a different concept and even when the drifted concept completely overlaps such space.

1.4 The approach taken

In this work, we analyze the psychophysical data presented above and consider it in the context of machine learning. We design a new protocol using unbalanced sample distribution for investigating the characteristics of the human visual system and its ability to cope with concept drift. The protocol is structured to help us better model the observed behavior and to challenge the morph effect limitations presented by Preminger et al. [39]. Using this protocol, we conduct a psychophysical experiment and collect additional data about the behavior of the human visual system.

Given the collective results, we devise here a strategy for machine learners to cope with concept drift without labels. The strategy includes an initial
phase in which a supervised learner is used to identify and label concepts. Once initial concepts are learned, an unsupervised learner slowly drifts concepts online. We name the unsupervised learner a Concept Follower (CF).

We adapt an unsupervised incremental learning algorithm known as the Leader Follower [7] 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 [29]. Two novel Concept Followers are presented and analyzed. The resulting algorithms offer the ability to adjust pre-learned concepts when facing gradually accumulating environmental changes. It is shown that by using unlabeled samples, a Concept Follower based on LF may handle concept drift even when the drifted concept completely overlaps a region previously learned as the center of a different concept. We provide experimental results as for the utility of the presented Concept Followers. The Concept Followers are tested with samples taken from the presentation protocol used by the psychophysical experiment and are shown to cope with a similar challenge. We demonstrate that the algorithms are stable and elastic.

1.5 Structure

The rest of the thesis is structured as follows: Chapter 2 presents and analyzes the psychophysical results collected as part of this research and com-
1.5. Structure

pares those to previous work. Chapter 3 summarizes our lessons learned from the visual system and introduces the framework of the machine learning study. Chapter 4 presents and analyzes the original Leader Follower algorithm and two novel Concept Follower algorithms. Last, chapter 5 describes an evaluation of the Concept Follower algorithms and summarizes our conclusions.
Chapter 2

Psychophysical Experiments

2.1 Overview

The electrophysiological and psychophysical reports presented above, suggest that unsupervised classification of spatially correlated images may add information to a representation. As a result, a category may gradually drift from its prelearned position in the stimulus continuum and expand. However, all reports correlated between similar objects, i.e. objects that are not easily distinguishable. For example Preminger et al. [39] intentionally chose faces which are not too different from each other. Note that even with such choice of face images, and following repeated exposure to the morph sequence, only half of the subjects experienced morph effect and expanded the category.

Such choice of stimuli and the results showing 50% of the subjects not experiencing morph effect, may lead the reader to suspect that the observed morph effect is not a robust phenomenon. As part of our research we collected additional experimental data about the human visual system, which allowed us to learn more about the biological abilities when coping with concept drift, and to model the observed behavior.
In the current study a novel protocol which arguably intensifies the morph effect is presented (See Hadas et al. [17]). A morphing sequence between significantly different face images was created to test the protocol. Using the new protocol, category drift was achieved following a single presentation of the morphing sequence and significantly faster than in previous reports. Former protocols, such as Preminger et al. [39], generated drifts of a category both away from and back towards the prelearned image due to the following reasons: (1) Since each session starts the morphed sequence from the prelearned object, some of the change in the perceptual boundary already achieved in the previous session is canceled. (2) Since every session presents the entire sequence of morphed images, subjects cross the categorization boundary between the representation of the prelearned image and an alternative representation on each session. Once the categorization boundary is crossed, the human visual system learns to discriminate between images of the morphed sequence rather than to generalize all images in the morphed sequence as belonging to the same representation.

In the experimental protocol presented here the drift of a category in the stimulus continuum is ensured to always accumulate away from the prelearned object by: (1) starting each new session with the intermediate morphed images that subjects classified as “targets” in the previous session, rather than going back to the first image in the morphed sequence; (2) presenting no more than one-third of the sequence of morphed images in every session to avoid crossing the category boundary. Using this novel paced mor-
phing experimental protocol, in just three days of exposure to a 10-minute session, subjects accumulated substantial drift and eventually classified a completely different image as the target.

2.2 Paced Morphing Experimental Setting

Following the introduction of a new protocol, in which morphing is paced, we evaluate the morph effect experienced by subjects. Our subjects learn a concept with feedback and then classify images without feedback. The images presented include morphed images adhering to the new protocol. We monitor the classification results and collect data about the response time and user experiences. We evaluate the morph effect robustness by using a morphing sequence between significantly different pre-learned concepts.

2.2.1 Stimuli

Faces were taken from a 100-face nottingham_scans image set (downloaded from http://pics.psych.stir.ac.uk/index.html). Eight faces were hand-picked from the database providing easy identification between them. Each of the eight faces was prepared by blacking out the background, hair, and ears. Two of the faces were named A and B and were used for the preparation of a sequence of morphed images. The constructed sequence included 99 morphed images. (See Figure 2.1) The sequence was prepared using Sqrinz
Chapter 2. Psychophysical Experiments

Figure 2.1: The Psychophysical Experiment Morphing Sequence
Images A and B and the sequence of morphed images prepared. The experiment presented 99 intermediate images with morphing ratio starting from 1:99 and ending at 99:1 between images B and A, respectively. Here 10 The remaining six images were picked as alternative images for the construction of the distracting face images. Ten morphed images were prepared from the first alternative image (inclusive) to the second (exclusive), ten more from the second image (inclusive) to the third (exclusive), etc. The resulting 50 alternative images were picked randomly whenever an image from the distracting face images was required.

Morph version 1.2e. Using a separate pair-wise similarity-rating test, it was confirmed that A and B are perceived as different from each other by naive observers. Observers evaluated 36 pairs of images, among which, the A and B pair appeared three times. On a scale in which 1 represents that the images are "exactly the same" and 5 represents that the images are "completely different", the A and B image pair received a mean rating of 4.09 (Std 1.04, n=11).

2.2.2 Apparatus

Images were presented using a 14-inch LCD of a laptop using 1024 x 768 resolution (refresh rate 60 Hz) in a small, quiet room containing fluorescent
lightening and no windows. Presented images were approximately 250 x 300 pixel Jpeg images on a black background. A fixation point was used when an image was not presented.

2.2.3 Subjects

Nine undergraduate students volunteered to participate in the experiment, to fulfill their introduction to psychology course requirements. Students were notified that they would be required to attend 30-min daily sessions on four different days. Students filled out a short questionnaire. Students were screened to ensure none were taking medications that might affect their memory or attention. Subjects were instructed that they would be required to learn a face and would be tested on how well they remembered the face during the experiment. The experiment was performed in accordance with the guidelines and regulations of the Ethics Committee of the Department of Psychology at Tel-Aviv University. Subjects’ consent was obtained prior to participation in the experiment.

2.2.4 Procedure

The experimental procedure is summarized in Figure 2.2 A. During all phases, subjects performed a target identification task - subjects were required to press the space bar in order to view a face and in this way controlled the presentation rate. Then, subjects were required to indicate whether they saw
Chapter 2. *Psychophysical Experiments*

the prelearned face by pressing 1 or an alternative face by pressing 2. The stimulus presentation lasted until the subject completed classification or 1000 ms poststimulus onset, whichever occurred first. Instructions were verbally provided by the instructor, on paper, and were also written on the bottom of the screen. During the target-learning phase, the upper screen was used to provide feedback to the subject. No feedback was provided in later phases.

Preliminary testing - Preliminary testing to obtain an adequate set of parameters for the protocol phase was employed. The parameters studied included the number of sessions, the interval used between sessions, the number of trials per session, the interval between images of the morphing sequence, and the exposure time. One lesson learned was that subjects became confused if the morphing sequence progressed too quickly. For example, using more than 33 % of the morphing sequence per day appeared to confuse many subjects. To avoid such confusion a presentation tool was programmed to limit the progress made per day to a range of 33 %.

Target-learning phase - In the target-learning phase, subjects learned image A as their target and were asked to remember it. Image A was presented for three seconds during three iterations. Subjects were asked to press key 1 on the keyboard after each presentation. Next, subjects started a target identification task with feedback using 20 trials. The images were chosen randomly using a 1:1 chance between image A and the distracting face images. The subject received feedback after classifying each image. At the end of this phase, the results were summarized and presented. A subject who
2.2. Paced Morphing Experimental Setting

performed this phase perfectly continued to the next phase. Eight of the nine subjects performed this phase perfectly while one subject made a single mistake and therefore repeated this phase.

Initial classification phase - Several minutes after the target-learning phase, subjects started the initial classification phase. In this phase the subjects performed 50 trials of a target identification task without feedback. The images were chosen randomly with a probability of 1:2 between image A and an alternative image. The alternative image was chosen with a probability of 1:4 between image B and the distracting face images. Note that image B was not presented prior to the initial classification phase, and the test did not provide an indication to allow subjects to single out image B from the distracting face images.

Protocol phase - A protocol phase immediately followed the initial classification phase without a break. Subjects were unaware of the transition between the tests and continued the target identification task without feedback as before. The presentation schedule during the protocol phase was chosen using the presentation tool and included a random 1:2 chance between images from the morph sequence and the distracting face images. Images A and B were not presented during this phase. The presentation tool was programmed to progress to the next image of the morphing sequence following a positive classification by a subject and to go back three images of the morphing sequence with every negative classification by a subject (staircase method). The protocol phase started on the first day and ended on the
fourth day when a subject completed the morphing sequence (reached 100 % morphing). In each daily session the experiment ended when the subject completed a continuous target identification task of 250 images (taking about 10 minutes). The first morphed image presented on each day was based on the last morphed image presented on the previous day. The presentation tool ensured that all morphed images presented during the session did not exceed a progression of 33 % morphing such that a minimum of four sessions are required to complete the morphing sequence.

Final classification phase - A final classification phase followed the protocol phase without a break. Subjects were unaware of the transition between the phases and continued the target identification task without feedback as before. During this phase, images were chosen randomly with 1:2 probability between image B and an alternative image. The alternative image was chosen with a 1:6 probability between image A and the distracting face images.

Debriefing - After each session, the subjects were asked to describe their experiences and indicate whether they had faced any difficulties during the task. Subject responses were documented.

2.3 Experiments and Results

The experimental results were published in Hadas et al. [17]. Nine subjects, unaware of the purpose of the experiment, learned to identify facial image A as their target and then undertook a “target identification task” by clas-
2.3. Experiments and Results

Classifying facial images as the same or different from the prelearned target. On average, when excluding the top and bottom 1% tail, classifications were done 803 ms (Std 232) after stimulus onset. The average resulting exposure time was 721 ms (Std 111). In just 3 days, during which a morphing sequence from face image A to a very different face image B was presented, eight of the nine subjects reverted to classifying image B as the prelearned target. This drastic change in representation occurred using just 210 ± 30 exposures to images from the morph sequence. All nine subjects showed similar performance during the first and second days of the experiment. On the third day, one subject drastically changed its classification patterns and reverted to classifying images from the morphing sequence as not belonging to the prelearned-target. Further, when presented with facial image A, the subject classified 94 % of facial image A trials as not belonging to the prelearned target. At this point his experiment was terminated.

During the protocol phase, eight subjects classified 98.7 % (Std 1.35 ) of the morphed faces as the prelearned face. Accordingly, the staircase method pushed these subjects along the morphing sequence away from image A and towards image B. Figure 2.2 B shows the average progress made by the eight subjects over the first three days. Note that all subjects made definite progress and accumulated considerable drift from image A. Apart from irregular lapses, drift accumulated steadily each day until the enforced daily limit of 33 % progress. The limit is expressed as a plateau at the end of each
day. On average, the eight subjects progressed during the three days from 0 % morphing to 97.25 % (Std 3.4 %). The protocol phase ended early on the fourth day when all subjects completed 100 % of morphing. The figure also shows the one subject (presented in dotted line) who drastically changed its classification patterns and reverted to classifying one of the distracting face images as the target. Following the critical stage, the subject classified morphed face images as not belonging to the target and at the same time, the subject has classified 93.75 % of the images presenting a particular distracting face image as belonging to the target (compared to 0 % prior to the critical stage).
2.3. Experiments and Results

Figure 2.2: The Psychophysical Experiment Layout

(A) Experimental layout. On day 1, subjects were trained to identify image A among nontarget distracter face images. Once successful, subjects performed an initial classification phase in which they discriminated image A from B and distracter faces. Then, unbeknown to the subject, the protocol phase began in which the A-to-B morphing sequence and distracters were presented, during which time, image A gradually transformed to image B in 1% morphing steps. During each day, subjects were limited to a maximum progression of 33% morphing. On the fourth day, subjects completed the classification of the entire morphing sequence. (B) Subject classification performance along the morphing sequence. Classifying a morphed image as belonging to the prelearned target face (face image A) was followed by a positive 1% progress in the morphing sequence. Classifying a morphed image as not belonging to the target was followed by a minus 3% change in the sequence. Apart from irregular lapses, eight subjects quickly progressed each day until the enforced daily limit of 33% progress (presented as a continuous line). The limit is expressed as a plateau at the end of each day. On average, subjects progressed during the three days from 0% morphing to 97.25% (Std 3.4). All eight subjects reached a morphing of 100% in the fourth day of the experiment (not shown). One subject (presented as a doted line) performed equally well compared to the other subjects until reaching a critical stage following which the subject had stopped referring to images from the morphing sequence as the target and her experiment was discontinued.
Subject classification patterns were affected by the morphing sequence progress as can be seen from Figure 2.3 (averaged over the eight subjects). A correlation of 0.87 was found between the spatial difference increase along the morph sequence and the percentage of morphed images classified as non-targets. The average percentage of morphed images classified as nontargets during the protocol phase was 3 % with Std of 3 % across the experiment days. The highest percentage of nontarget classifications was recorded during the last 10 % of the sequence with 9 % (Std 11 %) nontargets. In comparison, the average error rate of distracting face images during the protocol phase was relatively constant across the three training days (mean: 2.3; Std 0.2 %).

During the initial classification phase, all nine subjects classified 100 % of image A trials as the prelearned face. Image B and the distracting face images were classified as the target only 0.7 % (Std 2 %) and 2 % (Std 2 %) of the times respectively. These data suggest that all subjects clearly distinguished image B as not belonging to the prelearned target during that phase. By day 3 of the experiment, six of the subjects repeatedly classified the last image in the sequence, representing 99 % morphing as the prelearned face (100 % of the time), one subject classified the image representing 97 % morphing as the prelearned face 18 of 22 times, and another subject was indecisive regarding the classification of images above 90 % morphing, with just 68 % of trials classified as the prelearned face. As day 4 started, all eight subjects classified all remaining images up to 100 % morphing as belonging
2.3. Experiments and Results

Figure 2.3: The Psychophysical Experiment Target Classification Ratio
The ratio of images from the morphing sequence classified as the target. The ratio is presented per each 10% morphing range. Strikingly, subjects classified the 90-99% morphing range (which includes only 1-10% of the prelearned image A), as nontarget only 9% (Std 11%) of the trials.

to the target and thus ended the protocol phase.

Subject’s Experiences during Protocol Phase Overall upon debriefing, the eight subjects did not report confusion and indicated that they were confident in their classifications. Reports were made that the sessions became harder each day. The subject who referred to one of the distracting face images during the third day as the target did not report any difficulty during the first two days of the experiment. At the end of the third day, the subject had reported that during the session, she had suddenly realized that she is
classifying incorrectly and that the image she had learned two days ago was a different one than the image she was targeting. Following this discovery the subject had turned to classify the other image as the target.

Image A was initially presented during the initial classification phase and was classified as the target. During the protocol phase image A was not presented and subjects turned to classify image B as the target. Image A was presented again during the final classification phase. Did subjects merge image A with image B into one category?

Through the final classification phase we found that when faced with both A and B during the same session, subjects classified one as the target and the other as a non-target. Seven subjects maintained their ability to discriminate image A from image B and did not generalize the target representation to include both images. Out of the seven, three subjects classified image A as different from the target (99 % of trails Std 1.7) and kept on classifying image B as the prelearned face (98.6 % Std 2.3 % of trials). Four other subjects, when exposed to image A during the final classification phase, transposed their previous decision and from that point on classified the re-appearing image A as the target (91 % Std 6 % of trials). Image B was classified as different from the prelearned target (96 % Std 4 % of trials). One subject classified both A (100 %) and B (100 %) as the target and expressed no doubt in her classification decisions.
2.4 Discussion of the results

As described by Hadas et al. [17], all nine subjects evaluated, experienced a robust morph effect. In just three days of exposure to a 10-minute session, subjects accumulated substantial drift. Eight of nine subjects eventually classified a completely different image as the target. Our findings are consistent with prior reports [32, 39, 43], which demonstrated that the visual system may adapt object categories without an external supervisor. All reports collectively indicate that exposure to a morphing sequence between two object categories may result in all sequence images being classified the same. However, unlike previous reports, here it is shown that the said adaptation is rapid, is found in almost all participants and can be induced even between very different object categories.

In the current study subjects learned to identify face image A as their target and discriminate it from other face images including face image B. The representation of the target category therefore included information that allowed them to differentiate A from B. Yet, during the exposure to the morphing sequence, subjects continued to classify the entire sequence of morphed images as their target, even when the morph included less information from the prelearned image A and more information from a clearly different image B. This evidence is consistent with previous findings [32, 39, 43] and suggests that exposure to images spatially correlated to a prelearned object may change the object category representation. In contrast to previous re-
ports, it is demonstrated here that this process does not require that the morphed sequence will be repeated over many sessions. Instead, it is apparent that a single presentation of the ordered morphed sequence suffices to achieve a morph effect.

The finding presented here differ from those presented by Preminger et al. [39], although in both studies subjects were exposed to a protocol that included a presentation of morphing sequence during an unsupervised classification task. Here, all subjects exposed to the protocol experienced a morph effect and eight of nine completed the sequence within 3 days, whereas in Perminger et al. 2007, only 55% were reported to experience a morph effect and completing the sequence required more exposures and more time. The difference in the results may be explained by the different protocols used during the presentation of the morphing sequence. The protocol used here included two significant modifications to the protocol used by Perminger et al.: (1) The protocol used by Perminger starts from the beginning of the sequence in each daily session. The protocol used here starts each new session with the image that subjects classified as a “target” in the previous session to avoid cancellation of the drift achieved so far. (2) In each daily session the protocol used by Perminger stops at the end of the sequence after crossing the categorization boundary. Once the categorization boundary is crossed, the visual system is trained to discriminate between the morphed images. The protocol used here avoids crossing the category boundary by presenting no more than 33% of the morphing sequence per session. The intent behind the
two modifications was to allow drift of a category in the stimulus continuum to accumulate away from the prelearned object and avoid the accumulation of drift in the reverse direction toward the prelearned object.

Two major differences between the results presented here and those of Preminger et al. [39] suggest that the procedures used here intensified the morph effect. First, as mentioned above, all the subjects exposed to the novel protocol experienced a morph effect, whereas only 55% of the subjects exposed to the Preminger protocol did so. Second, the progress made by subjects exposed to the novel protocol appears to be significantly faster than progress made by subjects exposed to the Preminger protocol (3 days and 235 exposures vs. about 4-15 days and 400-1500 exposures). The new results add to the results of Preminger et al. [39] as it may suggest that morph effect is a robust phenomenon in the visual system rather than occurring in only about one out of two subjects, as may be suggested by the previous findings.

Another important difference between the experiment described here and the one described by Preminger et al. [39] arise when considering the preparation of the morphing sequence. In Preminger et al. [39], the sequence preparation ensured that there were no conspicuous features that may cause trivial discrimination between faces. The constructed faces were chosen to be neither too similar nor too different from each other. The choice of not too different faces while creating the morphing sequence may leave the reader wondering if a morph effect only occurs between fairly similar faces. Here, the face images were selected to be significantly different when creating the
sequence of morphed images. The results show that a morph effect also occurs using significantly different faces.

The results in Hadas et al. [17] suggest that the visual system is able to robustly cope with Concept Drift without a supervisor. Such an ability is demonstrated even when the concept drift is significant and in all subjects. The visual system copes with Concept Drift rapidly and using relatively small number of exposures. No parallel ability was demonstrated to-date in machine learners. Unsupervised machine learning methods for coping with Concept Drift may offer an advantage, compared to supervised methods, in problem areas where labels are scarce or not available in a timely manner.
Chapter 3

Computational Model

Characterization Based On The Psychophysical Findings

Traditional computational approach for coping with concept drift challenges require labeled data to adapt. As presented in Hadas et al. [17], biological systems do not require such labeled data to accommodate to change. After the Visual System learns new concepts with or without a supervisor, the system later adapt such concepts to environmental change without a supervisor. This biologically used strategy may offer a significant advantage over alternative strategies of using a supervisor to accommodate the change. Further, biological systems appear to cope well with a case in which a concept drifts into the space of a different concept. The behavior of biological systems is therefore of interest while considering machine algorithms for handling concept drift. In our research we seek to learn from this biological method.

Next, we summarize the psychophysical findings, serving as a basis for our computational approach. 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 unlabeled samples; (3) Unlabeled 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 toward 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 \[39\]. Here and in Hadas et al. \[17\], we presented results showing that the same strategy is used when the visual system was presented with unbalanced sample distribution in which only one concept was presented.

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

Next, we present a Concept Follower - an incremental unsupervised machine learning algorithm that may maintain a machine learner accuracy using unlabeled samples.
Chapter 3. Computational Model Characterization Based On The Psychophysical Findings
Chapter 4

Methodology

In this chapter, unsupervised Concept Followers are presented and analyzed. The strategy used includes an initial supervised stage. During the supervised stage, concept centers are learned from a multi-dimensional environment using labeled samples. Once concepts were learned, the presented Concept Followers use unlabeled samples to accommodate the concept centers 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 Concept Followers detect the abrupt change, to allow repeating the supervised stage.

The traditional Leader Follower algorithm (LF by Duda et al. [7]) 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 Concept Follower 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 unlabeled sample.
4.1 The traditional Leader Follower algorithm (LF)

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, 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. [7] algorithm, adjusting the concept towards the sample was achieved by adding the new sample to the concept and normalizing the resulting concepts at each step. A generalization of the Duda et al. [7] algorithm, which allowed any form of concept update towards the sample was presented by Garg and Shyamasundar [13]. This variant included learning without concept and sample normalization.
4.1. The traditional Leader Follower algorithm (LF)

- Initialize $\theta$ to the distance threshold between a sample and a concept

1: $s \leftarrow$ New sample
2: $N \leftarrow 1$
3: $w_N \leftarrow s$
4: loop
5: $s \leftarrow$ New sample
6: $j \leftarrow \arg\min_i \|s - w_i\|$ for any $i = 1..N$
7: if $\|s - w_j\| < \theta$ then
8: Add $s$ to $w_j$ and update $w_j$ towards $s$
9: else
10: $N \leftarrow N + 1$
11: $w_N \leftarrow s$
12: end if
13: end loop

The Garg and Shyamasundar [13] variant of the traditional Leader Follower algorithm is presented in Algorithm 1. The algorithm initializes 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 ($\theta$).

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

CF1 and CF2, presented next, include such modifications.

4.2 Concept Follower algorithm (CF1)

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 $\epsilon_0$. Then a Concept Follower algorithm (CF1), is used to drift the pre-learned concepts $\{w_i\}$. The concepts originally learned using older labeled samples are later adjusted based on newer unlabeled samples. This allows accommodating the concepts to changes in the environment.
4.2. Concept Follower algorithm (CF1)

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

1: $T \leftarrow 0$
2: $E \leftarrow 0$
3: loop
4: $T \leftarrow T + 1$
5: $s \leftarrow$ New sample
6: $j \leftarrow \text{argmin } \|s - w_i\|$ for any $i = 1..N$
7: if $\|s - w_j\| < \theta$ then
8: $w_j \leftarrow w_j + \eta \cdot (s - w_j)$
9: else
10: $E \leftarrow E + 1$
11: end if
12: if $T > T_{window}$ then
13: if $E/T > \epsilon_{max}$ then
14: Terminate and initiate new supervised learning
15: end if
16: $T \leftarrow 0$
17: $E \leftarrow 0$
18: end if
19: end loop

The Concept Follower algorithm (CF1) is derived from LF. Unlike LF, new concepts are not learned by CF1. Instead, samples that are not in the
Chapter 4. Methodology

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 ($\epsilon_{\text{max}}$), 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 ($\epsilon_{\text{max}}$) should be set higher then the error rate predicted during the initial supervised learning phase (i.e. $\epsilon_{\text{max}} > \epsilon_0$).

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 ($\theta$). If the distance is smaller than the threshold, the concept is considered a match and the concept is slightly shifted toward the classified sample to help correct to concept drift. The shift is controlled by a learning rate parameter ($\eta$). 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 toward a region previously learned as belonging
4.2. Concept Follower algorithm (CF1)

to a second concept. This may occur for example when the sample ratio of the second concept is 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.
Chapter 4. Methodology
4.3 Concept Follower with Unlearning algorithm (CF2)

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

1: $T \leftarrow 0$
2: $E \leftarrow 0$
3: loop
4: $T \leftarrow T + 1$
5: $s \leftarrow$ New sample
6: $j \leftarrow \arg\min_i \|s - w_i\|$ for any $i = 1..N$
7: if $\|s - w_j\| < \theta$ then
8: $w_j \leftarrow w_j + \eta \cdot (s - w_j)$
9: $w_i \leftarrow w_i - \delta \cdot (s - w_i)$ for all $i = 1..N; i \neq j; \|s - w_i\| < \theta$
10: else
11: $E \leftarrow E + 1$
12: end if
13: if $T > T_{window}$ then
14: if $E/T > \epsilon_{max}$ then
15: Terminate and initiate new supervised learning
16: end if
17: $T \leftarrow 0$
18: $E \leftarrow 0$
19: end if
20: end loop
Chapter 4. Methodology

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 toward 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.

When a new sample is processed, the distances of all concepts to the sample are calculated and compared with a predefined threshold parameter ($\theta$). 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 ($\epsilon_{\text{max}}$), 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
4.4 Mathematical Properties

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 toward the sample. The shift is controlled by a learning rate parameter ($\eta$). 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 ($\delta$). 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.

4.4 Mathematical Properties

We have introduced two algorithms that are capable of adjusting without labels a set of concepts to environmental changes. We next show that under certain assumptions, the presented algorithms CF1 and CF2 converges. Through such convergence, the distance between a learned concept and its respective concept center in the problem area decreases over time. As a re-
sult of this convergence, when the concept center drifts in the problem area, the learned concept will be adapted toward the concept center.

CF1 and CF2 adapt a learned concept toward its samples. We show that repeatedly adapting a learned concept toward its samples results in decreasing the distance between the learned concept and the concept center as present in the problem area.

![Figure 4.1: Expected adaptation of a learned concept](image)

An example monotonic spherical PDF of a concept center is shown and an example learned concept is presented. Since the learned concept is located at distance $d$ from the concept center, the expected adaptation of the learned concept should drift the learned concept toward the concept center.

We assume that the concept probability density function (PDF) is monotonic, spherical and can be approximated by observing the Euclidean distance between a sample and the learned concept center (See Figure 4.1). Using this
4.4. Mathematical Properties

assumption we suggest to always adapt the concept toward its samples such that the adaptation is proportional to the distance between the sample and the concept center.

We mark the direction from the learned concept toward the concept center as the positive direction. Given a sample \( (s) \) and a concept \( (w_j) \), the concept adaptation \( (\Delta) \) used by CF1 and CF2 is proportional to the distance between the concept and the sample and is controlled by the learning rate parameter \( (\eta) \).

\[
\Delta = \eta \cdot (s - w_j)
\]

We mark the concept PDF as \( f(x) \) where \( x \) is the distance between the concept center and a sample. We further mark the distance between the learned concept \( w_j \) and the concept center as \( d \) such that the distance between the learned concept and the sample is \( (x + d) \). We need to show that the expectancy of \( \Delta \) is both in the direction of the concept center (i.e. positive) and is smaller then \( d \) (such that the learned concept converges to the concept center).

\[
0 < E[\Delta] < d
\]

The expected adaptation can be written as:

\[
E[\Delta] = E[\eta \cdot (x + d)] = \eta \cdot E[(x + d)]
\]

\[
E[\Delta] = \eta \cdot \int (x + d) f(x) \, dx = \eta \int x f(x) \, dx + d\eta \int f(x) \, dx
\]
But note that the concept is assumed to be spherical and therefore its expectancy is zero.

\[ \int x f(x) \, dx = 0 \]

And that

\[ \int f(x) \, dx = 1 \]

Choosing \( 0 < \eta < 1 \) will result in

\[ E[\Delta] = d\eta > 0 \]

We have shown that the expectancy of \( \Delta \) is both positive and is smaller then \( d \). The expected adaptation of the learned concept is to reduce its distance from the concept center in the problem area. We note that \( E[\Delta] = 0 \) iff \( d = 0 \) and \( E[\Delta] > 0 \) iff \( d > 0 \). In other words, the learned concept is expected to stay in place if it is located where the concept center in the problem area is and will adapt toward the center otherwise.

We have presented CF1 and CF2 algorithms that adapt learned concepts toward the concept center in the problem area. Such adaptation is proportional to the distance between the learned concept and the sample and is controlled by a learning rate parameter. Under the assumption that concept PDFs are of monotonic and spherical, we have shown that CF1 and CF2 drift a learned concept toward the respective concept center in the problem area. This convergence property of CF1 and CF2 ensures that the learned
4.4. Mathematical Properties

concept is adapted toward the concept center in the problem area if sampled. CF1 and CF2 convergence allow dynamic and yet stable adaptation of the learned concept during a concept drift.

In CF1, we suggested to adjust only the concept closest to the sample. When a concept drifts, CF1 uses unlabeled samples classified as belonging to that concept to adapt the concept center. However if the concept is not sampled, CF1 has no indication of its drift. In order to support problems areas with unbalanced sample ratio between concepts where a concept may drift while not being sampled, we introduce CF2. It is suggested that in some problem areas it may be beneficial to adjust all concepts in the proximity of the sample.

The following chapters describe our computational experiments and the results obtained by CF1 and CF2. We compared between the results of CF1 and CF2. We also compared the results of our unsupervised CF1 and CF2 to the results of a simple supervised incremental learner, which maintains a sliding window with the ten latest labeled samples per each concept. The ten labels are used by the supervised incremental learner to calculate the concept center.
Chapter 5

Computational Experiments

5.1 Experimental Setup

In order to evaluate the behavior of the CF1 and CF2 algorithms, we used samples taken from a morphing sequence between two face image exemplars. 101 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. [17]. 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
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 to 200.

from the morphing sequence were centered 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 5.1). Each of the 101 low resolution images was represented using a 551 dimensional vector to depict the $(29 \times 19 =) 551$ 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 to 200 such that the morphing sequence starts with sample 100 representing the first original face and gradually continues toward 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 centers were programmatically set to the two face image exemplars. The unsupervised algorithm concepts were initialized 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 5.2). The assessment process after each sample is shown in Table 5.1. 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 center of 'A'. The sample that was closest to Concept 'B', following each sample learned was named the concept center of 'B'.
The Sequential Presentation Protocol (2A) and the Random Presentation Protocol (2B) included interlaced assessment periods following each processed sample.

\[
SP_{i,k} \leftarrow \frac{1}{1 + \|s_i - w_k\|} \\
CC_k \leftarrow \arg\min_i \|s_i - w_k\|
\]

Table 5.1: The assessment process.

\(s_i\) (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 \(k\).

Figure 5.2: The presentation protocols used

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...
5.2. Results

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. At first, CF2 was initialized 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 (See Figure 5.3).

![Figure 5.3: Two shapes morphing sequence.](image)

A Square is morphed into a Circle. Here only 5 out of 100 morphed shapes are shown.

### 5.2 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 proto-
col, both CF1 and CF2 followed Concept 'A' as it gradually drifted toward 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.

5.2.1 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 ten latest labels of each concept were equally weighted and used to determine the concept center. We tested under the assumption that not all samples are labeled. Figure 5.4 shows the results when using 10%, 2.5% and 0% of samples being labeled. Note that substantial adaptation of the concept was exhibited with 10% of samples being labeled. 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 labeled modulates the adaptation of Concept 'A'.
5.2. Results

Figure 5.4: Final Sample Proximity supervised learner.
The three plots show the results of a sliding window incremental supervised learner that use an average of the ten latest labels per concept. The right plot shows the results when 10% of the samples are labeled. 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 labeled. 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 labeled. Concept 'A' and Concept 'B' roughly divide the morphing space. Concept 'A' had not shifted towards Concept 'B'.

5.2.2 Concept Follower algorithm (CF1)

The Concept Follower algorithm (CF1) was used with a variety of learning rate ($\eta$) and threshold ($\theta$) parameters. Figure 5.5 shows one exemplar of the results for a case where the learning rate parameter is $\eta = 0.1$. The threshold ($\theta$) 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
Figure 5.5: Final Sample Proximity using CF1. Results for CF1 with $\eta = 0.1$, $\theta = 100$. 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.

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 initialized. 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
5.2. Results

Figure 5.6: Concept center drift results using CF1. Results for CF1 with $\eta = 0.1$, $\theta = 100$. The right subplot shows the drift of Concept 'A' center during the presentation of 101 samples representing a morphing sequence from sample 100 and sample 200; Concept 'A' gradually drifts toward Concept 'B'. Once the two Concepts overlap, Concept 'A' discontinues its drift. The left subplot shows the drift of concept centers 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'. Samples at the range 143-200 are closer to Concept 'B'.

Figure 5.6 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 center. The right plot shows the concept centers 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 toward the center and later back toward 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 centers during the random presentation. It can be seen that Concepts 'A' and 'B' quickly drifts toward the center 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.
5.2. Results

5.2.3 Concept Follower with Unlearning algorithm (CF2)

Figure 5.7: Final Sample Proximity using CF2.
Results for CF2 with $\eta = 0.1$, $\delta = 0.05$, $\theta = 100$. 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.

The Concept Follower with Unlearning algorithm (CF2) was used with a variety of learning rate ($\eta$), unlearning rate ($\delta$) and threshold ($\theta$) parameters. Figure 5.7 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 ($\theta$) 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 mor-
Chapter 5. Computational Experiments

Phing 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 initialized. 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 5.8 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 center. The right plot shows the concept centers 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 centers during the random presentation. It can be seen that Concepts 'A' and 'B' quickly drift toward the center 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
5.2. Results

Figure 5.8: Concept center drift results using CF2.
Results for CF2 with $\eta = 0.1$, $\delta = 0.05$, $\theta = 100$. The right subplot shows the drift of Concept 'A' center during the presentation of 101 samples representing a morphing sequence from sample 100 and sample 200; Concept 'A' gradually drifts toward Concept 'B'. Once the two Concepts overlap, Concept 'A' continues its drift toward sample 200 and Concept 'B' drifts away from the morphing sequence. The left subplot shows the drift of concept centers 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.

200 to approximately two equal halves.

5.2.4 Sensitivity to The Learning Rate Parameter

The sensitivity of the presented protocols to the learning rate parameter was evaluated. In this experiment, CF1 was repeatedly used with different learning rate ($\eta$) parameters. We started with $\eta = 0.001$ and then factored $\eta$ by 1.2 per each iteration till we reached $\eta = 0.85$. All results are presented on a logarithmic scale. The threshold ($\theta$) was set to 100.
Chapter 5. Computational Experiments

Figure 5.9: The effect of the learning rate parameter on CF1. Results for CF1 with $\eta = 0.001...0.85$, $\theta = 100$. 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 centers 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 centers.

Figure 5.9 shows the effect of the learning rate parameter ($\eta$) on the results. The left plot 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. It is shown that no errors appear with learning rates below 0.5. The middle plot shows the standard deviation of the concepts when facing a Random Presentation Protocol (the measurements are taken along the second 500 random samples in each iteration). It is shown that increasing the learning rate increases the standard deviation suggesting greater concept instability. The right plot shows the final location of the concepts following a Sequential Presentation Protocol. Note that given the entire range of learning rates,
5.2. Results

the concept behavior is qualitatively the same. Concept 'A' drifts more than halfway towards Concept 'B' for learning rates greater than 0.05.

5.2.5 Sensitivity to The Threshold Parameter

The sensitivity of the presented protocols to the threshold parameter was evaluated. In this experiment, CF1 was repeatedly used with different threshold (θ) parameters. We started with θ = 1 (under a 551 dimensional space in which the difference between a white pixel and a black pixel is 255) and then factored θ by 1.2 per each iteration till we reached θ = 1021. All results are presented on a logarithmic scale. The learning rate (η) was set to 0.1.

Figure 5.10: The effect of the threshold parameter on CF1. Results for CF1 with θ = 1…1021, η = 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 centers 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 < θ < 1000 allows us to avoid errors, offer an approximately constant stability and reach the same final setup of concept centers.

Figure 5.10 shows the effect of the threshold parameter (θ) on the results.
Chapter 5. Computational Experiments

The left plot 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. It is shown that errors appear with thresholds below 80. The middle plot shows the standard deviation of the concepts when facing a Random Presentation Protocol (the measurements are taken along the second 500 random samples in each iteration). Notice that no specific pattern appears with thresholds above 80. The right plot shows the final location of the concepts following a Sequential Presentation Protocol. Note that using Threshold above 30, the concept behavior is identical.

5.2.6 Sensitivity to The Unlearning Rate Parameter

The sensitivity of the presented protocols to the unlearning rate parameter was evaluated. In this experiment, CF2 was repeatedly used with different unlearning rate ($\delta$) parameters. We started with $\delta = 0.001$ and then factored $\delta$ by 1.2 per each iteration till we reached $\delta = 0.7$. All results are presented on a logarithmic scale. The learning rate ($\eta$) was set to 0.1 while the threshold ($\theta$) was set to 100.

Figure 5.11 shows the effect of the unlearning rate parameter ($\delta$) on the results. The left plot 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.
5.2. Results

Figure 5.11: The effect of the unlearning rate parameter on CF2. Results for CF2 with $\delta = 0.001...0.7$, $\eta = 0.1$, $\theta = 100$. 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 centers 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 centers.

It is shown that no errors appear with unlearning rates below or equal to the learning rate (0.1). The middle plot shows the standard deviation of the concepts when facing a Random Presentation Protocol (the measurements are taken along the second 500 random samples in each iteration). It is shown that increasing the unlearning rate above the learning rate (0.1) significantly increases the standard deviation, suggesting greater concept instability. When setting the unlearning rates below the learning rate (0.1), the unlearning rate does not appear to modulate the concept stability. The right plot shows the final location of the concepts following a Sequential Presentation Protocol. Note that given unlearning rates below 0.005, Concept 'B' is not unlearned, while choosing unlearning rate of 0.005 or above appears
to be insignificant to the results.

5.2.7 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 ($\theta$) was set to 1000 (under a 1600 dimensional space in which the difference between a white pixel and a black pixel is 255).

Figure 5.12, top row shows the eight concepts after CF2 was initialized 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.
5.3 Discussion

Figure 5.12: Shape morphing.
Eight concepts of CF2 with $\eta = 0.2$, $\theta = 1000$, $\delta = 0.01$, were initialized 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.

5.3 Discussion

This research was motivated by a psychophysically observed strategy in which unlabeled samples were successfully used to incrementally adapt pre-learned concepts while coping with dramatic environmental changes. Two unsupervised incremental Concept Follower algorithms (CF1 and CF2) were develop-
The presented Concept Followers were adapted from the Leader Follower algorithm (LF). 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 behavior 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 Concept Followers 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 labeled 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 [17, 39, 43]. 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 centers. It is further assumed that initial concept centers can be determined by a supervised learning stage. Such limitations of CF1 and CF2 may be lifted by developing Concept Followers based on other unsupervised incremental learning algorithms.

Unsupervised Concept Followers are useful in problem areas in which the concept behavior 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 characterized by having a sharp PDF boundary area where the probability of samples inside the area is significantly higher then 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 labeled samples and later adapt such concepts using unlabeled samples [17, 32, 39, 43]. Classifying unlabeled stimulus affects future classifications of resembling stimuli. The system adapts incre-
mentaly 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 unlabeled samples and affect future classifications of stimuli resembling the classified stimulus. If the environment change continues, 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 [17]. We tested CF1 and CF2 under similar conditions. Although CF1 is able to drift the first concept and accumulate change toward 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
5.3. Discussion

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. 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 have introduced an incremental algorithm to follow changing pre-learned concepts without labeled 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 labeled 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
Chapter 5. Computational Experiments

leads to a constant accuracy penalty. Unsupervised Concept Followers help tune a machine learner shortly after a change, leading to a smaller accuracy penalty after each change. Depending on the availability of labeled samples, unsupervised Concept Followers 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, labeled samples are only available for some of the samples but not for all. When changes occur more frequently than the rate in which new labeled samples become available, accumulated changes may go undetected and when the next labeled sample appears it may represent an abrupt change. Depending on the supervised method used, abrupt changes may not allow incremental learning [29]. The frequency of labeled samples and of changes may therefore, limit the suitability of certain incremental supervised learning methods. Unsupervised Concept Followers 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 labeled samples are scarce, supervised learners face a greater challenge of identifying true environmental changes from noise. Unsupervised Concept Followers utilize all samples for learning and therefore may offer better robustness to noise under certain problem areas.

Unsupervised Concept Followers 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 la-
beled samples to evaluate the correctness of unsupervised Concept Followers. Alternatively, labeled samples can be used in combination with unlabeled samples to drift the learner concepts [See semi-supervised learning, for example, 48]. Another disadvantage of the presented unsupervised Concept Followers, 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.

We have demonstrated the ability of the Concept Followers to adjust to gradual change and qualitatively compared it to the performance exhibited by the human visual system. We therefore used a dataset 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 Concept Followers 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 [4], 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 Concept Followers. 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 Concept Follower algorithms. It is shown that the learning and unlearning have a local effect and can therefore be used in a multi-concept environment.

5.4 Conclusions and Future Work

We presented CF1 and CF2, unsupervised incremental learning algorithms based on the Leader Follower algorithm, for adapting concepts in a changing
5.4. Conclusions and Future Work

environment. It was argued that depending on the problem area, the use of unsupervised Concept Followers 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.

The dataset chosen to evaluate the machine learner 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 datasets 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. Additionally, more 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.
Bibliography


