ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8. Available from http://www.i6doc.com/en/.

Incremental learning algorithms and applications

Alexander Gepperth[1], Barbara Hammer[2] *

1- UIIS, ENSTA ParisTech
INRIA, Université Paris-Saclay
828 Bvd des Maréchaux, 91762 Palaiseau Cedex, France
2- Bielefeld University, CITEC centre of excellence
Universitätsstrasse 21-23
D-33594 Bielefeld, Germany

Abstract. Incremental learning refers to learning from streaming data, which arrive over time, with limited memory resources and, ideally, without sacrificing model accuracy. This setting fits different application scenarios such as learning in changing environments, model personalisation, or lifelong learning, and it offers an elegant scheme for big data processing by means of its sequential treatment. In this contribution, we formalise the concept of incremental learning, we discuss particular challenges which arise in this setting, and we give an overview about popular approaches, its theoretical foundations, and applications which emerged in the last years.

1 What is incremental learning?

Machine learning methods offer particularly powerful technologies to infer structural information from given digital data; still, the majority of current applications restrict to the classical batch setting: data are given prior to training, hence meta-parameter optimisation and model selection can be based on the full data set, and training can rely on the assumption that the data and its underlying structure are static. *Incremental learning*, in contrast, refers to the situation of continuous model adaptation based on a constantly arriving data stream [38, 149]. This setting is present whenever systems act autonomously such as in autonomous robotics or driving [5, 65, 112, 156]. Further, online learning becomes necessary in interactive scenarios where training examples are provided based on human feedback over time [134]. Finally, many digital data sets, albeit static, can become so big that they are de facto dealt with as a data stream, i.e. one incremental pass over the full data set [116]. Incremental learning investigates how to learn in such streaming settings. It comes in various forms in the literature, and the use of the term is not always consistent. Therefore, first, we give a meaning to the relevant terms *online learning*, *incremental learning*, and *concept drift*, giving particular attention to the supervised learning paradigm.

1.1 Online learning methods

In supervised learning, data $\mathcal{D} = ((\vec{x}_1, y_1), (\vec{x}_2, y_2), (\vec{x}_3, y_3), \ldots, (\vec{x}_m, y_m))$ are available with input signals $\vec{x}_i$ and outputs y_i. The task is to infer a model

*This research/work was supported by the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG). Alexander Gepperth is also with INRIA FLOWERS.

$M \approx p(y|\vec{x})$ from such data. Machine learning algorithms are often trained in a *batch* mode, i.e., they use all examples $(\vec{x}_i, y_i)$ at the same time, irrespective of their (temporal) order, to perform, e.g., a model optimisation step.

Challenge 1: Online model parameter adaptation. In many application examples, data $\mathcal{D}$ are not available priorly, but examples arrive over time, and the task is to infer a reliable model M_t after *every* time step based on the example $(\vec{x}_t, y_t)$ and the previous model M_{t-1} only. This is realised by *online learning* approaches, which use training samples one by one, without knowing their number in advance, to optimise their internal cost function. There is a continuum of possibilities here, ranging from fully online approaches that adapt their internal model immediately upon processing of a single sample, over so-called *mini-batch* techniques that accumulate a small number of samples, to batch learning approaches, which store all samples internally.

Online learning is easily achieved by stochastic optimisation techniques such as online back-propagation, but there are also extensions of the support vector machine (SVM) [164]. Prototype-based models such as vector quantisation, radial basis function networks (RBF), supervised learning vector quantisation (LVQ), and self-organising maps (SOM) all naturally realise online learning schemes, since they rely on an (approximate) stochastic gradient technique [15, 83, 115, 140]. Second order numeric optimisation methods and advanced optimisation schemes can be extended as well, such as variational Bayes, convex optimization, second order perceptron learning based on higher order statistics in primal or dual space, and online realisations of the quasi-Newton Broyden-Fletcher-Goldfarb-Shanno technique [49, 62, 114, 117, 125]. Stochastic optimization schemes can be developed also for non-decomposable cost function, [80]. Further, lazy learners such as k-nearest neighbour (k-NN) methods lend itself to online scenarios by their design [140]. Interestingly, online learning has already very early been accompanied by exact mathematical investigations [162].

1.2 Incremental learning methods

Incremental learning refers to online learning strategies which work with *limited memory resources*. This rules out approaches which essentially work in batch mode for the inference of M_t by storing all examples up to time step t in memory; rather, incremental learning has to rely on a compact representation of the already observed signals, such as an efficient statistics of the data, an alternative compact memory model, or an implicit data representation in terms of the model parameters itself. At the same time, it has to provide accurate results for all relevant settings, despite its limited memory resources.

Challenge 2: Concept drift. Incremental learning shares quite a number of challenges with online learning, with memory limitations adding quite a few extras. One prominent problem consists in the fact that, when the temporal structure of data samples is taken into account, one can observe changes in data statistics that occur over time, i.e. samples $(\vec{x}_i, y_i)$ are not i.i.d. Changes in the data distribution over time are commonly referred to as *concept drift* [33, 88, 126, 157]. Different types of concept drift can be distinguished: changes in the input distribution $p(\vec{x})$ only, referred to as *virtual concept drift* or *covariate shift*, or changes in the underlying functionality itself $p(y|\vec{x})$, referred to as *real concept drift*. Further, concept drift can be gradual or abrupt. In the former case one often uses the term *concept shift*. The term *local concept drift* characterises changes of the data statistics only in a specific region of the data space [157]. A prominent example is the addition of a new, visually dissimilar object class to a classification problem. Real concept drift is problematic since it leads to conflicts in the classification, for example when a new but visually similar class appears in the data: this will in any event have an impact on classification performance until the model can be re-adapted accordingly.

Challenge 3: The stability-plasticity dilemma. In particular for noisy environments or concept drift, a second challenge consists in the question when and how to adapt the current model. A quick update enables a rapid adaptation according to new information, but old information is forgotten equally quickly. On the other hand, adaption can be performed slowly, in which case old information is retained longer but the reactivity of the system is decreased. The dilemma behind this trade-off is usually denoted the *stability-plasticity dilemma*, which is a well-known constraint for artificial as well as biological learning systems [113]. Incremental learning techniques, which adapt learned models to concept drift only in those regions of the data space where concept drift actually occurs, offer a partial remedy to this problem. Many online learning methods alone, albeit dealing with limited resources, are not able to solve this dilemma since they exhibit a so-called *catastrophic forgetting* behaviour [44, 45, 108, 103, 132] even when the new data statistics do not invalidate the old ones.

One approach to deal with the stability-plasticity dilemma consists in the enhancement of the learning rules by explicit meta-strategies, when and how to learn. This is at the core of popular incremental models such as ART networks [56, 77], or meta-strategies to deal with concept drift such as the just-in-time classifier JIT [3], or hybrid online/offline methods [43, 120]. One major ingredient of such strategies consists in a confidence estimation of the actual model prediction, such as statistical tests, efficient surrogates, or some notion of self-evaluation [8, 43, 78]. Such techniques can be enhanced to complex incremental schemes for interactive learning or learning scaffolding [84, 130].

Challenge 4: Adaptive model complexity and meta-parameters. For incremental learning, model complexity must be variable, since it is impossible to estimate the model complexity in advance if the data are unknown. Depending on the occurrence of concept drift events, an increased model complexity might become necessary. On the other hand, the overall model complexity is usually bounded from above by the limitation of the available resources. This requires the intelligent reallocation of resources whenever this limit is reached. Quite a number of approaches propose intelligent adaptation methods for the model complexity such as incremental architectures [166], self-adjustment of the number of basic units in extreme learning machines [31, 177] or prototype-based models [77, 98, 144], incremental base function selection for a sufficiently powerful data representation [23], or self-adjusting cluster numbers in unsupervised learning [79]. Such strategies can be put into the more general context of self-evolving systems, see e.g. [92] for an overview. An incremental model complexity is not only mandatory whenever concept drift is observed, hence a possibly changing model complexity is present, but it can also dramatically speed-up learning in batch scenarios, since it makes often tedious model selection superfluous.

In batch learning, not only the model complexity, but also essential meta-parameters such as learning rate and strength of regularisation are determined prior to training. Often, time consuming cross-validation is used in batch learning, whereby first promising results how to automate this process exist [155]. However, these are not suited for incremental learning scenarios: Concept drift turns critical meta-parameters such as the learning rate into model parameters, since their choice has to be adapted according to the (changing) data characteristics. Due to this fact, incremental techniques often rely on models with few and robust meta-parameters (such as ensembles), or they use meta-heuristics how to adapt these quantities during training.

Challenge 5: Efficient memory models. Due to their limited resources, incremental learning models have to store the information provided by the observed data in compact form. This can be done via suitable system invariants (such as the classification error for explicit drift detection models [33]), via the model parameters in implicit form (such as prototypes for distance- based models [63]), or via an explicit memory model [96, 98]. Some machine learning models offer a seamless transfer of model parameters and memory models, such as prototype- or exemplar–based models, which store the information in the form of typical examples [63]. Explicit memory models can rely on a finite window of characteristic training examples, or represent the memory in the form of a parametric model. For both settings, a careful design of the memory adaptation is crucial since it directly mirrors the stability-plasticity dilemma [96, 98].

Challenge 6: Model benchmarking. There exist two fundamentally different possibilities to assess the performance of incremental learning algorithms:
(1) Incremental -vs- non-incremental: In particular in the absence of concept drift, the aim of learning consists in the inference of the stationary distribution $p(y|\vec{x})$ for typical data characterised by $p(\vec{x})$. This setting occurs e.g. whenever incremental algorithms are used for big data sets, where they compete with often parallelized batch algorithms. In such settings, the method of choice evaluate the classification accuracy of the final model M_t on a test set, or within a cross-validation. While incremental learning should attain results in the same range as batch variants, one must take into account that they deal with restricted knowledge due to their streaming data access. It has been shown, as an example, that incremental clustering algorithms cannot reach the same accuracy as batch versions if restricted in terms of their resources [2].
(2) Incremental -vs- incremental: When facing concept drift, different cost functions can be of interest. Virtual concept drift aims for the inference of a stationary model $p(y|\vec{x})$ with drifting probability $p(\vec{x})$ of the inputs. In such settings, the robustness of the model when evaluated on test data which follow a possibly skewed distribution is of interest. Such settings can easily be generated e.g. by enforcing imbalanced label distributions for test and training data [73]. Whenever real confidence drift is present, the online behaviour of the classification error $\|M_t(\vec{x}_{t+1}) - y_{t+1}\|$ for the next data point is usually the method of choice; thereby, a simple average of these errors can be accompanied by a de-

tailed inspection of the overall shape of the online error, since it provides insight into the rates of convergence e.g. for abrupt concept drift.

(3) Formal guarantees on the generalisation behaviour: Since many classical algorithms such as the simple perceptron or large margin methods have been proposed as online algorithms, there exists an extensive body of work investigating their learning behaviour, convergence speed, and generalisation ability, classically relying on the assumption of data being i.i.d. [162]. Some results weaken the i.i.d. assumption e.g. requiring only interchangeability [146]. Recently, popular settings such as learning a (generalised) linear regression could be accompanied by convergence guarantees for arbitrary distributions $p(\vec{x})$ by taking a game theoretic point of view: in such settings, classifier M_t and training example $\vec{x}_{t+1}$ can be taken in an adversial manner, still allowing fast convergence rates in relevant situations [87, 131, 151, 158]. The approach [117] even provides first theoretical results for real context drift, i.e. not only the input distribution, but also the conditional distribution $p(y|\vec{x})$ can follow mild changes.

2 Incremental learning models

Incremental learning comes in various forms in the literature, and the use of the term is not always consistent; for some settings, as an example, a memory limitation cannot be guaranteed, or models are designed for stationary distributions only. We will give an overview over popular models in this context. Thereby, we will mostly focus on supervised methods due to its popularity. Online or incremental learning techniques have also been developed for alternative tasks such as clustering [91, 109], dimensionality reduction [6, 12, 24, 25, 93, 123], feature selection and data representation[42, 27, 59, 72, 173, 179], reinforcement learning [11, 60], mining and inference [54, 129].

Explicit treatment of concept drift. Dealing with concept drift at execution time constitutes a challenging task [33, 88, 126, 157]. There exist different techniques to address concept drift, depending on its type. Mere concept shift is often addressed by so-called passive methods, i.e. learning technologies which smoothly adapt model parameters such that the current distribution is reliably represented by the model. Rapid concept changes, however, often require active methods, which detect concept drift and react accordingly.

Virtual concept drift, which concerns the input distribution only, can easily occur e.g. due to highly imbalanced classes over time. One popular state-of-the-art technology accounts for this fact by so-called importance weighting, i.e. strategies which explicitly or implicitly re-weight the observed samples such that a greater robustness is achieved [10, 73, 81]. Alternatively, concept shift can have its reason in novelty within the data or even new classes. Such settings can naturally be incorporated into local models provided they offer an adaptive model complexity [43, 56, 100, 133, 144].

Real concept drift can be detected by its effect on characteristic features of the model such as the classification accuracy. Such quantitative features can be accompanied by statistical tests which can judge the significance of their chance, hence concept drift. Tests can rely on well-known statistics such as the Hoeffding bound [48], or alternatively on suitable distances such as the Hellinger distance,

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8. Available from http://www.i6doc.com/en/.

which can measure the characteristics of value distributions of such characteristic features. When integrated into robust classifiers such as ensemble techniques, models which can simultaneously deal with different types of drift can result [16].

Support vector machines and generalised linear models. Several incremental SVM models exist [164]. Some rely on heuristics, like retraining a model with all support vectors plus a new "incremental" batch of data [35, 152], but without theoretical guarantees. Other incorporate modification of the SVM cost function to facilitate incrementality [141] and also possibly control complexity [58, 57]. Still, their resources are not strictly limited. As an alternative, *adiabatically* SVM training has been proposed, i.e., presenting one example at a time while maintaining the relevant optimality conditions on all previously seen examples. However this requires all previously seen samples to be stored, although the approach can considerably simplify SVM training. Ensemble learning algorithms based on SVM [127, 164] achieve incremental learning by training new classifiers for new batches of data, and combining all existing classifiers only for decision making. Another hybrid scheme combines a SVM classifier with a prototype-based data representation, whereby the latter can be designed as an online model based on which training examples for SVM can be generated [169]. Alternatively, SVMs can directly be trained in primal space, where online learning is immediate [22]. Online versions have also been proposed for other generalised linear models such as Gaussian Process regression [53, 110], whereby none of these models can yet easily deal with concept drift.

Connectionist models. As the problem of catastrophic forgetting was first remarked for multilayer perceptrons (MLP) [108, 132], it is hardly surprising that there exists significant work how to avoid it in connectionist systems. Initial consensus traced catastrophic forgetting back to their *distributed* information representation [46]. Indeed, *localist* connectionist models such as RBF networks can work reliably in incremental settings [100, 133], whereby care has to be taken to guarantee their generalisation performance [147]. Both capabilities are combined in *semi-distributed representations.* A number of algorithmic modifications of the MLP model has been proposed, such as sparsification [45], orthogonalization of internal node weights [47, 119], reduction of representational overlap while training [85], or specific regularisation [55]. These are successful in mitigating but not eliminating catastrophic forgetting [147]. Recently, there has been an increased interest in extreme learning machines (ELM), which combine a random mapping with a trained linear readout. Due to their simple training, incremental variants can easily be formulated, whereby their reservoir naturally represents rich potential concepts [31, 61, 159, 178].

Furthermore, there exist attempts to modify the system design of MLPs [86, 150] which are more in the line of generative learning; they incorporate novelty detection and use different representational resources for new samples. Elaborate connectionist models feature different memory subsystems for long- and short-term learning [7, 139], as well as explicit replay and re-learning of previous samples to alleviate forgetting [135]. These approaches reduce the problem of catastrophic forgetting at the price of vastly more complex model. Contrarily to other modern approaches, inspiration is taken primarily from biology and thus

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8. Available from http://www.i6doc.com/en/.

its solid mathematical understanding is yet lacking.

Explicit partitioning approaches. Many modern incremental learners rely on a local partitioning of the input space, and a separate classification/regression model for each partition [18, 21, 121, 148, 160]. The manner of performing this partitioning is very diverse, ranging from kd-trees [21] to genetic algorithms [18] and adaptive Gaussian receptive fields [160]. Equally, the choice of local models varies between linear models [160], Gaussian mixture regression [21] or Gaussian Processes [121]. For high-dimensional problems such as occur in perception, the partitioning of the input space constitutes the bottleneck as concerns memory consumption. Covariance matrices as used in [160], for example, are quadratic in the number of input dimensions, hence prohibitive for high dimensional data.

Decision trees partially alleviate this problem insofar as they cut along one dimension for every branching only, disregarding feature correlations. Quite a number of incremental tree builders have been proposed for classification [41, 52, 142], with a particular focus on when to split, how to avoid overly large trees while incremental growth, and how to reliably deal with imbalanced classes [26, 66, 102]. Interestingly, there do exist tree classifiers which result is entirely invariant to the ordering of the training data, but at the price of unlimited resources [90].

Ensemble methods. Ensemble methods combine a collection of different models by a suitable weighting strategy. As such, they are ideally suited to implicitly represent even partially contradictory concepts in parallel and mediate the current output according to the observed data statistics at hand. Ensemble methods have proved particularly useful when dealing with concept drift, with a few popular models ranging from incremental random forests [105], ensembles of bipartite graph classifiers [13], up to advanced weighting schemes suitable for different types of concept drift and recurring concepts [32, 39, 95, 111, 172].

Prototype-based methods. Prototype-based machine learning has its counterpart in cognitive psychology [137] which hypothesises that semantic categories in the human mind are represented by specific examples for these categories. In machine learning approaches, a class is represented by a number of representatives, and class membership is defined based on the distance of the data from these prototypes. For high dimensional data, adaptive low-rank metric learning schemes can dramatically improve classification accuracy and efficiency [17, 145]. Prototype-based methods are a natural continuation of the work on localist or semi-distributed representations in early connectionist models, and thus share many properties. They have the advantage of an easily adaptive model complexity. One disadvantage is that the number of prototypes can become large whenever complex class boundaries are present.

Prototype-based models are closely connected to the non-parametric k-NN classifier (all training points act as prototypes) and the RBF model [140]. A popular supervised method is given by LVQ and recent variants which can be substantiated by a cost function [15]. A number of incremental variants and methods capable of dealing with concept drift have been proposed, such as dynamic prototype inversion / deletion schemes [98, 144], or techniques with fixed model complexity, but intelligent source redistribution strategies [50]. Similar unsupervised incremental models exist [19, 63, 176].

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8. Available from http://www.i6doc.com/en/.

Insights into biological incremental learning. As biological incremental learning has reached a high degree of perfection, biological paradigms can provide inspiration how to set up artificial incremental systems. There is evidence that sensory representations in the neocortex are prototype-based, whereby neurons are topologically arranged by similarity [40, 94, 138, 153]. Learning acts on these representations in a task-specific way insofar as the density of neurons is correlated to sensory regions which require finer discrimination [128], i.e., where more errors occur. Here, learning is conceivably enhanced through acetylcholine release in case of task failures [70, 163]. Learning respects the topological layout by changing only a small subset of neural selectivities [136] at each learning event, corresponding to regions around the best matching unit [40].

Beyond the single-neuron level, there is a large body of literature investigating the roles of the hippocampal and neocortical areas of the brain in learning at the architectural level. Generally speaking, the hippocampus employs a rapid learning rate with separated representations whereas the neocortex learns slowly, building overlapping representations of the learned task [122]. A well-established model of the interplay between the hippocampus and the neocortex suggests that recent memories are first stored in the hippocampal system and played back to the neocortex over time [107]. This accommodates the execution of new tasks that have not been recently performed as well as the transfer of new task representations from the hippocampus (short-term memory) to the neocortical areas (long-term memory) through slow synaptic changes, i.e. it provides an architecture which is capable of facing the stability-plasticity dilemma.

3 Applications

We would like to conclude this overview by a glimpse on typical application scenarios where incremental learning plays a major role.

Data analytics and big data processing. There is an increasing interest in single-pass limited-memory models which enable a treatment of big data within a streaming setting [64]. The aim is to reach the capability of offline techniques, hence conditions are less strict as concerns e.g. the presence of concept drift. Recent approaches extend, for example, extreme learning machines in this way [168]. Domains, where this approach is taken, include image processing [34, 97], data visualisation [106], and processing of networked data [29].

Robotics. Autonomous robotics and human-machine-interaction are inherently incremental, since they are open-ended, and data arrive as a stream of signals with possibly strong drift. Incremental learning paradigms have been designed in the realm of autonomous control [161], service robotics [5], computer vision [175], self-localisation [82], or interactive kinesthetic teaching [51, 143]. Further, the domain of autonomous driving is gaining enormous speed [4, 118, 156], with enacted autonomous vehicle legislation in already eight states in the US (Dec. 2015). Another emerging area, caused by ubiquitous sensors within smart phones, addresses activity recognition and modeling [1, 68, 69, 74, 89, 99].

Image processing. Image and video data are often gathered in a streaming fashion, lending itself to incremental learning. Typical problems in this context range from object recognition [9, 36, 98], image segmentation [36, 71], and image representation [30, 165], up to video surveillance, person identification, and visual tracking [28, 37, 101, 104, 134, 154, 167, 174].

Automated annotation. One important process consists in the automated annotation or tagging of digital data. This requires incremental learning approaches as soon as data arrive over time; example systems are presented in the approaches [14, 20, 75] for video and speech tagging.

Outlier detection. Automated surveillance of technical systems equipped with sensors constitutes an important task in different domains, starting from process monitoring [67], fault diagnosis in technical systems [76, 170, 171], up to cyber-security [124]. Typically, a strong drift is present in such settings, hence there is a high demand for advanced incremental learning techniques.

References

[1] Z. Abdallah, M. Gaber, B. Srinivasan, and S. Krishnaswamy. Adaptive mobile activity recognition system with evolving data streams. *Neurocomputing*, 150(PA):304–317, 2015.

[2] M. Ackerman and S. Dasgupta. Incremental clustering: The case for extra clusters. In *NIPS*, pages 307–315, 2014.

[3] C. Alippi, G. Boracchi, and M. Roveri. Just in time classifiers: Managing the slow drift case. In *IJCNN*, pages 114–120, 2009.

[4] R. Allamaraju, H. Kingravi, A. Axelrod, G. Chowdhary, R. Grande, J. How, C. Crick, and W. Sheng. Human aware UAS path planning in urban environments using nonstationary MDPs. In *IEEE International Conference on Robotics and Automation*, pages 1161–1167, 2014.

[5] Y. Amirat, D. Daney, S. Mohammed, A. Spalanzani, A. Chibani, and O. Simonin. Assistance and service robotics in a human environment. *Robotics and Autonomous Systems*, 75, Part A:1 – 3, 2016.

[6] A. Anak Joseph and S. Ozawa. A fast incremental kernel principal component analysis for data streams. In *IJCNN*, pages 3135–3142, 2014.

[7] B. Ans and S. Rousset. Avoiding catastrophic forgetting by coupling two reverberating neural networks. *Academie des Sciences, Sciences de la vie*, 320, 1997.

[8] S.-H. Bae and K.-J. Yoon. Robust online multi-object tracking based on tracklet confidence and online discriminative appearance learning. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1218–1225, 2014.

[9] X. Bai, P. Ren, H. Zhang, and J. Zhou. An incremental structured part model for object recognition. *Neurocomputing*, 154:189–199, 2015.

[10] A. Balzi, F. Yger, and M. Sugiyama. Importance-weighted covariance estimation for robust common spatial pattern. *Pattern Recognition Letters*, 68:139–145, 2015.

[11] A. Barreto, D. Precup, and J. Pineau. On-line reinforcement learning using incremental kernel-based stochastic factorization. In *NIPS*, volume 2, pages 1484–1492, 2012.

[12] N. Bassiou and C. Kotropoulos. Online PLSA: Batch updating techniques including out-of-vocabulary words. *IEEE Transactions on Neural Networks and Learning Systems*, 25(11):1953–1966, 2014.

[13] J. Bertini, M. Do Carmo Nicoletti, and L. Zhao. Ensemble of complete P-partite graph classifiers for non-stationary environments. In *CEC*, pages 1802–1809, 2013.

[14] S. Bianco, G. Ciocca, P. Napoletano, and R. Schettini. An interactive tool for manual, semi-automatic and automatic video annotation. *Computer Vision and Image Understanding*, 131:88–99, 2015.

[15] M. Biehl, A. Ghosh, and B. Hammer. Dynamics and generalization ability of LVQ algorithms. *Journal of Machine Learning Research*, 8, 2007.

[16] D. Brzezinski and J. Stefanowski. Reacting to different types of concept drift: The accuracy updated ensemble algorithm. *IEEE Transactions on Neural Networks and Learning Systems*, 25(1):81–94, 2014.

[17] K. Bunte, P. Schneider, B. Hammer, F. Schleif, T. Villmann, and M. Biehl. Limited rank matrix learning, discriminative dimension reduction and visualization. *Neural Networks*, 26:159–173, 2012.

[18] M. Butz, D. Goldberg, and P. Lanzi. Computational complexity of the xcs classifier system. *Foundations of Learning Classifier Systems*, 51, 2005.

[19] Q. Cai, H. He, and H. Man. Imbalanced evolving self-organizing learning. *Neurocomputing*, 133:258–270, 2014.

[20] H. Carneiro, F. França, and P. Lima. Multilingual part-of-speech tagging with weightless neural networks. *Neural Networks*, 66:11–21, 2015.

[21] T. Cederborg, M. Li, A. Baranes, and P.-Y. Oudeyer. Incremental local online gaussian mixture regression for imitation learning of multiple tasks. 2010.

[22] O. Chapelle. Training a support vector machine in the primal. *Neural Comput.*, 19(5):1155–1178, May 2007.

[23] H. Chen, P. Tino, and X. Yao. Efficient probabilistic classification vector machine with incremental basis function selection. *IEEE Transactions on Neural Networks and Learning Systems*, 25(2):356–369, 2014.

[24] Y. Choi, S. Ozawa, and M. Lee. Incremental two-dimensional kernel principal component analysis. *Neurocomputing*, 134:280–288, 2014.

[25] K. Cui, Q. Gao, H. Zhang, X. Gao, and D. Xie. Merging model-based two-dimensional principal component analysis. *Neurocomputing*, 168:1198–1206, 2015.

[26] R. De Rosa and N. Cesa-Bianchi. Splitting with confidence in decision trees with application to stream mining. In *IJCNN*, volume 2015-September, 2015.

[27] A. Degeest, M. Verleysen, and B. Frénay. Feature ranking in changing environments where new features are introduced. In *IJCNN*, volume 2015-September, 2015.

[28] M. Dewan, E. Granger, G.-L. Marcialis, R. Sabourin, and F. Roli. Adaptive appearance model tracking for still-to-video face recognition. *Pattern Recognition*, 49:129–151, 2016.

[29] C. Dhanjal, R. Gaudel, and S. Clémençon. Efficient eigen-updating for spectral graph clustering. *Neurocomputing*, 131:440–452, 2014.

[30] K. Diaz-Chito, F. Ferri, and W. Diaz-Villanueva. Incremental generalized discriminative common vectors for image classification. *IEEE Transactions on Neural Networks and Learning Systems*, 26(8):1761–1775, 2015.

[31] J.-L. Ding, F. Wang, H. Sun, and L. Shang. Improved incremental regularized extreme learning machine algorithm and its application in two-motor decoupling control. *Neurocomputing*, (Part A):215–223, 2015.

[32] G. Ditzler and R. Polikar. Incremental learning of concept drift from streaming imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 25(10):2283–2301, 2013.

[33] G. Ditzler, M. Roveri, C. Alippi, and R. Polikar. Learning in nonstationary environments: A survey. *IEEE Computational Intelligence Magazine*, 10(4):12–25, 2015.

[34] T.-N. Doan, T.-N. Do, and F. Poulet. Parallel incremental svm for classifying million images with very high-dimensional signatures into thousand classes. In *IJCNN*, 2013.

[35] C. Domeniconi and D. Gunopulos. Incremental support vector machine construction. In *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, pages 589–592, 2001.

[36] J. Dou, J. Li, Q. Qin, and Z. Tu. Moving object detection based on incremental learning low rank representation and spatial constraint. *Neurocomputing*, 168:382–400, 2015.
[37] J. Dou, J. Li, Q. Qin, and Z. Tu. Robust visual tracking based on incremental discriminative projective non-negative matrix factorization. *Neurocomputing*, 166:210–228, 2015.
[38] E. Eaton, editor. *Lifelong Machine Learning, AAAI Spring Symposium*, volume SS-13-05 of *AAAI Technical Report*. AAAI, 2013.
[39] R. Elwell and R. Polikar. Incremental learning of concept drift in nonstationary environments. *IEEE Transactions on Neural Networks*, 22(10):1517–1531, 2011.
[40] C. A. Erickson, B. Jagadeesh, and R. Desimone. Clustering of perirhinal neurons with similar properties following visual experience in adult monkeys. *Nature neuroscience*, 3(11):1143–1148, 2000.
[41] J. Fan, J. Zhang, K. Mei, J. Peng, and L. Gao. Cost-sensitive learning of hierarchical tree classifiers for large-scale image classification and novel category detection. *Pattern Recognition*, 48(5):1673–1687, 2015.
[42] A. Ferreira and M. Figueiredo. Incremental filter and wrapper approaches for feature discretization. *Neurocomputing*, 123:60–74, 2014.
[43] L. Fischer, B. Hammer, and H. Wersing. Combining offline and online classifiers for life-long learning. In *IJCNN*, volume 2015-September, 2015.
[44] R. French. Connectionist models of recognition memory: constraints imposed by learning and forgetting functions. *Psychol Rev.*, 97(2), 1990.
[45] R. French. Semi-distributed representations and catastrophic forgetting in connectionist networks. *Connect. Sci.*, 4, 1992.
[46] R. French. Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences*, 3(4), 1999.
[47] R. M. French. Dynamically constraining connectionist networks to produce distributed, orthogonal representations to reduce catastrophic interference. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*. 1994.
[48] I. Frias-Blanco, J. Del Campo-Ávila, G. Ramos-Jiménez, R. Morales-Bueno, A. Ortiz-Díaz, and Y. Caballero-Mota. Online and non-parametric drift detection methods based on Hoeffding's bounds. *IEEE Transactions on Knowledge and Data Engineering*, 27(3):810–823, 2015.
[49] C. Gentile, F. Vitale, and C. Brotto. On higher-order perceptron algorithms. In *NIPS*, pages 521–528, 2007.
[50] A. Gepperth and C. Karaoguz. A bio-inspired incremental learning architecture for applied perceptual problems. *Cognitive Computation*, 2015. accepted.
[51] A. Ghalamzan E., C. Paxton, G. Hager, and L. Bascetta. An incremental approach to learning generalizable robot tasks from human demonstration. In *ICRA*, volume 2015-June, pages 5616–5621, 2015.
[52] A. Gholipour, M. Hosseini, and H. Beigy. An adaptive regression tree for non-stationary data streams. In *Proceedings of the ACM Symposium on Applied Computing*, pages 815–816, 2013.
[53] A. Gijsberts and G. Metta. Real-time model learning using incremental sparse spectrum gaussian process regression. *Neural Networks*, 41:59–69, 2013.
[54] J. Gomes, M. Gaber, P. Sousa, and E. Menasalvas. Mining recurring concepts in a dynamic feature space. *IEEE Transactions on Neural Networks and Learning Systems*, 25(1):95–110, 2014.
[55] I. J. Goodfellow, M. Mirza, X. Da, A. Courville, and Y. Bengio. An empirical investigation of catastrophic forgeting in gradient-based neural networks. *arXiv preprint arXiv:1312.6211*, 2013.
[56] S. Grossberg. Adaptive resonance theory: How a brain learns to consciously attend, learn, and recognize a changing world. *Neural Networks*, 37:1–47, 2013.
[57] B. Gu, V. Sheng, K. Tay, W. Romano, and S. Li. Incremental support vector learning for ordinal regression. *IEEE Transactions on Neural Networks and Learning Systems*, 26(7):1403–1416, 2015.
[58] B. Gu, V. Sheng, Z. Wang, D. Ho, S. Osman, and S. Li. Incremental learning for ν-support vector regression. *Neural Networks*, 67:140–150, 2015.
[59] N. Guan, D. Tao, Z. Luo, and B. Yuan. Online nonnegative matrix factorization with robust stochastic approximation. *IEEE Transactions on Neural Networks and Learning Systems*, 23(7):1087–1099, 2012.
[60] P. Guan, M. Raginsky, and R. Willett. From minimax value to low-regret algorithms for online markov decision processes. In *Proceedings of the American Control Conference*, pages 471–476, 2014.
[61] L. Guo, J.-H. Hao, and M. Liu. An incremental extreme learning machine for online sequential learning problems. *Neurocomputing*, 128:50–58, 2014.
[62] E. Hall and R. Willett. Online convex optimization in dynamic environments. *IEEE Journal on Selected Topics in Signal Processing*, 9(4):647–662, 2015.
[63] B. Hammer and A. Hasenfuss. Topographic mapping of large dissimilarity datasets. *Neural Computation*, 22(9):2229–2284, 2010.
[64] B. Hammer, H. He, and T. Martinetz. Learning and modeling big data. In M. Verleysen, editor, *ESANN*, pages 343–352, 2014.
[65] B. Hammer and M. Toussaint. Special issue on autonomous learning. *KI*, 29(4):323–327, 2015.
[66] A. Hapfelmeier, B. Pfahringer, and S. Kramer. Pruning incremental linear model trees with approximate lookahead. *IEEE Transactions on Knowledge and Data Engineering*, 26(8):2072–2076, 2014.
[67] L. Hartert and M. Sayed-Mouchaweh. Dynamic supervised classification method for online monitoring in non-stationary environments. *Neurocomputing*, 126:118–131, 2014.
[68] M. Hasan and A. Roy-Chowdhury. Incremental activity modeling and recognition in streaming videos. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 796–803, 2014.
[69] M. Hasan and A. Roy-Chowdhury. Incremental learning of human activity models from videos. *Computer Vision and Image Understanding*, 144:24–35, 2016.
[70] M. E. Hasselmo. The role of acetylcholine in learning and memory. *Current opinion in neurobiology*, 16(6):710–715, 2006.
[71] J. He, L. Balzano, and A. Szlam. Incremental gradient on the Grassmannian for online foreground and background separation in subsampled video. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1568–1575, 2012.
[72] X. He, P. Beauseroy, and A. Smolarz. Dynamic feature subspaces selection for decision in a nonstationary environment. *International Journal of Pattern Recognition and Artificial Intelligence*, 2015.
[73] T. Hoens and N. Chawla. Learning in non-stationary environments with class imbalance. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 168–176, 2012.
[74] W. Hu, X. Li, G. Tian, S. Maybank, and Z. Zhang. An incremental dpmm-based method for trajectory clustering, modeling, and retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(5):1051–1065, 2013.
[75] L. Huang, X. Liu, B. Ma, and B. Lang. Online semi-supervised annotation via proxy-based local consistency propagation. *Neurocomputing*, 149(PC):1573–1586, 2015.
[76] S.-Y. Huang, F. Yu, R.-H. Tsaih, and Y. Huang. Network-traffic anomaly detection with incremental majority learning. In *IJCNN*, volume 2015-September, 2015.
[77] S. Impedovo, F. Mangini, and D. Barbuzzi. A novel prototype generation technique for handwriting digit recognition. *Pattern Recognition*, 47(3):1002–1010, 2014.
[78] A. Jauffret, C. Grand, N. Cuperlier, P. Gaussier, and P. Tarroux. How can a robot evaluate its own behavior? a neural model for self-assessment. In *IJCNN*, 2013.
[79] A. Kalogeratos and A. Likas. Dip-means: An incremental clustering method for estimating the number of clusters. In *NIPS*, volume 3, pages 2393–2401, 2012.
[80] P. Kar, H. Narasimhan, and P. Jain. Online and stochastic gradient methods for non-decomposable loss functions. In *NIPS*, volume 1, pages 694–702, 2014.
[81] H. Kawakubo, M. C. du Plessis, and M. Sugiyama. Computationally efficient class-prior estimation under class balance change using energy distance. *IEICE Transactions*, 99-D(1):176–186, 2016.
[82] S. Khan and D. Wollherr. Ibuild: Incremental bag of binary words for appearance based loop closure detection. In *ICRA*, volume 2015-June, pages 5441–5447, 2015.
[83] T. Kohonen. Self-organized formation of topologically correct feature maps. *Biol. Cybernet.*, 43:59–69, 1982.
[84] V. Kompella, M. Stollenga, M. Luciw, and J. Schmidhuber. Explore to see, learn to perceive, get the actions for free: Skillability. In *IJCNN*, pages 2705–2712, 2014.
[85] C. Kortge. Episodic memory in connectionist networks. In *Proceedings of the 12th Annual Conference of the Cognitive Science Society*. 1990.

[86] J. Krushke. ALCOVE: An exemplar-based model of category learning. *Psychological Review*, 99, 1992.
[87] E. Kuhn, J. Kolodziej, and R. Seara. Analysis of the tdlms algorithm operating in a nonstationary environment. *Digital Signal Processing: A Review Journal*, 45:69–83, 2015.
[88] P. Kulkarni and R. Ade. Incremental learning from unbalanced data with concept class, concept drift and missing features: a review. *International Journal of Data Mining and Knowledge Management Process*, 4(6), 2014.
[89] I. Kviatkovsky, E. Rivlin, and I. Shimshoni. Online action recognition using covariance of shape and motion. *Computer Vision and Image Understanding*, 129:15–26, 2014.
[90] B. Lakshminarayanan, D. Roy, and Y. Teh. Mondrian forests: Efficient online random forests. In *NIPS*, volume 4, pages 3140–3148, 2014.
[91] R. Langone, O. Mauricio Agudelo, B. De Moor, and J. Suykens. Incremental kernel spectral clustering for online learning of non-stationary data. *Neurocomputing*, 139:246–260, 2014.
[92] A. Lemos, W. Caminhas, and F. Gomide. Evolving intelligent systems: Methods, algorithms and applications. *Smart Innovation, Systems and Technologies*, 13:117–159, 2013.
[93] Y. Leng, L. Zhang, and J. Yang. Locally linear embedding algorithm based on omp for incremental learning. In *IJCNN*, pages 3100–3107, 2014.
[94] D. A. Leopold, I. V. Bondar, and M. A. Giese. Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature*, 442(7102):572–575, 2006.
[95] P. Li, X. b. Wu, X. Hu, and H. Wang. Learning concept-drifting data streams with random ensemble decision trees. *Neurocomputing*, 166:68–83, 2015.
[96] D. Liu, M. Cong, Y. Du, and X. Han. Robotic cognitive behavior control based on biology-inspired episodic memory. In *ICRA*, volume 2015-June, pages 5054–5060, 2015.
[97] L. Liu, X. Bai, H. Zhang, J. Zhou, and W. Tang. Describing and learning of related parts based on latent structural model in big data. *Neurocomputing*, 173:355–363, 2016.
[98] V. Losing, B. Hammer, and H. Wersing. Interactive online learning for obstacle classification on a mobile robot. In *IJCNN*, volume 2015-September, 2015.
[99] C. Loy, T. Xiang, and S. Gong. Incremental activity modeling in multiple disjoint cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(9):1799–1813, 2012.
[100] J. Lu, F. Shen, and J. Zhao. Using self-organizing incremental neural network (soinn) for radial basis function networks. In *IJCNN*, pages 2142–2148, 2014.
[101] Y. Lu, K. Boukharouba, J. Boonært, A. Fleury, and S. Lecœuche. Application of an incremental svm algorithm for on-line human recognition from video surveillance using texture and color features. *Neurocomputing*, 126:132–140, 2014.
[102] R. Lyon, J. Brooke, J. Knowles, and B. Stappers. Hellinger distance trees for imbalanced streams. In *ICPR*, pages 1969–1974, 2014.
[103] M. McCloskey and N.J. Cohen. Catastrophic interference in connectionist networks: the sequential learning problem. *Psychol. Learn. Motiv.*, 24, 1989.
[104] C. Ma and C. Liu. Two dimensional hashing for visual tracking. *Computer Vision and Image Understanding*, 135:83–94, 2015.
[105] K. Ma and J. Ben-Arie. Compound exemplar based object detection by incremental random forest. In *ICPR*, pages 2407–2412, 2014.
[106] Z. Malik, A. Hussain, and J. Wu. An online generalized eigenvalue version of laplacian eigenmaps for visual big data. *Neurocomputing*, 173:127–136, 2016.
[107] J. L. McClelland, B. L. McNaughton, and R. C. O'Reilly. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102:419–457, 1995.
[108] M. McCloskey and N. Cohen. Catastrophic interference in connectionist networks: the sequential learning problem. In G. H. Bower, editor, *The psychology of learning and motivation*, volume 24. 1989.
[109] S. Mehrkanoon, O. Agudelo, and J. Suykens. Incremental multi-class semi-supervised clustering regularized by kalman filtering. *Neural Networks*, 71:88–104, 2015.
[110] F. Meier, P. Hennig, and S. Schaal. Incremental local Gaussian regression. In *NIPS*, volume 2, pages 972–980, 2014.
[111] D. Mejri, R. Khanchel, and M. Limam. An ensemble method for concept drift in nonstationary environment. *Journal of Statistical Computation and Simulation*, 83(6):1115–1128, 2013.
[112] E. Menegatti, K. Berns, N. Michael, and H. Yamaguchi. Special issue on intelligent autonomous systems. *Robotics and Autonomous Systems*, 74, Part B:297 – 298, 2015. Intelligent Autonomous Systems (IAS-13).
[113] M. Mermillod, A. Bugaiska, and P. Bonin. The stability-plasticity dilemma: investigating the continuum from catastrophic forgetting to age-limited learning effects. *Frontiers in Psychology*, 4:504, 2013.
[114] A. Mokhtari and A. Ribeiro. Global convergence of online limited memory BFGS. *Journal of Machine Learning Research*, 16:3151–3181, 2015.
[115] J. Moody and C. J. Darken. Fast learning in networks of locally tuned processing units. *Neural Computation*, 1, 1989.
[116] G. D. F. Morales and A. Bifet. Samoa: Scalable advanced massive online analysis. *Journal of Machine Learning Research*, 16:149–153, 2015.
[117] E. Moroshko, N. Vaits, and K. Crammer. Second-order non-stationary online learning for regression. *Journal of Machine Learning Research*, 16:1481–1517, 2015.
[118] A. Mozaffari, M. Vajedi, and N. Azad. A robust safety-oriented autonomous cruise control scheme for electric vehicles based on model predictive control and online sequential extreme learning machine with a hyper-level fault tolerance-based supervisor. *Neurocomputing*, 151(P2):845–856, 2015.
[119] J. Murre. The effects of pattern presentation on interference in backpropagation networks. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society*. 1992.
[120] Q. Nguyen and M. Milgram. Combining online and offline learning for tracking a talking face in video. In *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops 2009*, pages 1401–1408, 2009.
[121] D. Nguyen-Tuong and J. Peters. Local gaussian processes regression for real-time model-based robot control. In *IEEE/RSJ International Conference on Intelligent Robot Systems*, 2008.
[122] R. C. OŔeilly. The division of labor between the neocortex and hippocampus. *Connectionist Models in Cognitive Psychology*, page 143, 2004.
[123] S. Ozawa, Y. Kawashima, S. Pang, and N. Kasabov. Adaptive incremental principal component analysis in nonstationary online learning environments. In *IJCNN*, pages 2394–2400, 2009.
[124] S. Pang, Y. Peng, T. Ban, D. Inoue, and A. Sarrafzadeh. A federated network online network traffics analysis engine for cybersecurity. In *IJCNN*, volume 2015-September, 2015.
[125] A. Penalver and F. Escolano. Entropy-based incremental variational bayes learning of gaussian mixtures. *IEEE Transactions on Neural Networks and Learning Systems*, 23(3):534–540, 2012.
[126] R. Polikar and C. Alippi. Guest editorial learning in nonstationary and evolving environments. *IEEE Transactions on Neural Networks and Learning Systems*, 25(1):9–11, 2014.
[127] R. Polikar, L. Upda, S. S. Upda, and V. Honavar. Learn++: An incremental learning algorithm for supervised neural networks. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 31(4):497–508, 2001.
[128] D. B. Polley, E. E. Steinberg, and M. M. Merzenich. Perceptual learning directs auditory cortical map reorganization through top-down influences. *The journal of neuroscience*, 26(18):4970–4982, 2006.
[129] M. Pratama, S. Anavatti, P. Angelov, and E. Lughofer. Panfis: A novel incremental learning machine. *IEEE Transactions on Neural Networks and Learning Systems*, 25(1):55–68, 2014.
[130] M. Pratama, J. Lu, S. Anavatti, E. Lughofer, and C.-P. Lim. An incremental meta-cognitive-based scaffolding fuzzy neural network. *Neurocomputing*, 171:89–105, 2016.
[131] A. Rakhlin, K. Sridharan, and A. Tewari. Online learning via sequential complexities. *Journal of Machine Learning Research*, 16:155–186, 2015.
[132] R. Ratcliff. Connectionist models of recognition memory: constraints imposed by learning and forgetting functions. *Psychological Review*, 97, 1990.
[133] P. Reiner and B. Wilamowski. Efficient incremental construction of rbf networks using quasi-gradient method. *Neurocomputing*, 150(PB):349–356, 2015.
[134] J. Rico-Juan and J. Iñesta. Adaptive training set reduction for nearest neighbor classification. *Neurocomputing*,

138:316–324, 2014.

[135] A. Robins. Catastrophic forgetting, rehearsal, and pseudorehearsal. *Connection Science*, 7, 1995.

[136] E. T. Rolls, G. Baylis, M. Hasselmo, and V. Nalwa. The effect of learning on the face selective responses of neurons in the cortex in the superior temporal sulcus of the monkey. *Experimental Brain Research*, 76(1):153–164, 1989.

[137] E. Rosch. Cognitive reference points. *Cognitive Psychology*, 7, 1975.

[138] D. A. Ross, M. Deroche, and T. J. Palmeri. Not just the norm: Exemplar-based models also predict face aftereffects. *Psychonomic bulletin & review*, 21(1):47–70, 2014.

[139] J. Rueckl. Jumpnet: A multiple-memory connectionist architecture. In *In Proceedings of the 15th Annual Conference of the Cognitive Science Society*. 1993.

[140] T. A. Runkler. *Data Analytics Models and Algorithms for Intelligent Data Analysis*. Springer Vieweg, 2012.

[141] S. Ruping. Incremental learning with support vector machines. In *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, pages 641–642, 2001.

[142] C. Salperwyck and V. Lemaire. Incremental decision tree based on order statistics. In *IJCNN*, 2013.

[143] M. Saveriano, S.-I. An, and D. Lee. Incremental kinesthetic teaching of end-effector and null-space motion primitives. In *ICRA*, volume 2015-June, pages 3570–3575, 2015.

[144] F.-M. Schleif, X. Zhu, and B. Hammer. Sparse conformal prediction for dissimilarity data. *Annals of Mathematics and Artificial Intelligence (AMAI)*, 74(1-2):95–116, 2015.

[145] P. Schneider, M. Biehl, and B. Hammer. Adaptive relevance matrices in learning vector quantization. *Neural Computation*, 21(12):3532–3561, 2009.

[146] G. Shafer and V. Vovk. A tutorial on conformal prediction. *JMLR*, 9:371–421, 2008.

[147] N. Sharkey and A. Sharkey. An analysis of catastrophic interference. *Connection Science*, 7(3-4), 1995.

[148] O. Sigaud, C. Sagaün, and V. Padois. On-line regression algorithms for learning mechanical models of robots: A survey. *Robotics and Autonomous Systems*, 2011.

[149] D. L. Silver. Machine lifelong learning: Challenges and benefits for artificial general intelligence. In *Artificial General Intelligence - 4th International Conference, AGI 2011*, pages 370–375, 2011.

[150] S. Sloman and D. Rumelhart. Reducing interference in distributed memories through episodic gating. In A. Healy and S. K. R. Shiffrin, editors, *Essays in Honor of W. K. Estes*. 1992.

[151] M. Sugiyama, M. Yamada, and M. C. du Plessis. Learning under nonstationarity: covariate shift and class-balance change. *Wiley Interdisciplinary Reviews: Computational Statistics*, 5(6):465–477, 2013.

[152] N. A. Syed, S. Huan, L. Kah, and K. Sung. Incremental learning with support vector machines. In *Proceedings of the Workshop on Support Vector Machines at the International Joint Conference on Articial Intelligence (IJCAI-99)*, 1999.

[153] K. Tanaka. Inferotemporal cortex and object vision. *Annual review of neuroscience*, 19(1):109–139, 1996.

[154] L. Tao, S. Mein, W. Quan, and B. Matuszewski. Recursive non-rigid structure from motion with online learned shape prior. *Computer Vision and Image Understanding*, 117(10):1287–1298, 2013.

[155] C. Thornton, F. Hutter, H. H. Hoos, and K. Leyton-Brown. Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms. In *Proc. of KDD-2013*, pages 847–855, 2013.

[156] S. Thrun. Toward robotic cars. *Commun. ACM*, 53(4):99–106, Apr. 2010.

[157] A. Tsymbal. The problem of concept drift: definitions and related work. Technical report, Computer Science Department, Trinity College Dublin, 2004.

[158] T. van Erven, P. D. Grünwald, N. A. Mehta, M. D. Reid, and R. C. Williamson. Fast rates in statistical and online learning. *Journal of Machine Learning Research*, 16:1793–1861, 2015.

[159] A. van Schaik and J. Tapson. Online and adaptive pseudoinverse solutions for elm weights. *Neurocomputing*, (Part A):233–238, 2015.

[160] S. Vijayakumar and S. Schaal. Locally weighted projection regression: An o(n) algorithm for incremental real time learning in high-dimensional spaces. In *International Conference on Machine Learning*, 2000.

[161] M. Wang and C. Wang. Learning from adaptive neural dynamic surface control of strict-feedback systems. *IEEE Transactions on Neural Networks and Learning Systems*, 26(6):1247–1259, 2015.

[162] T. L. H. Watkin, A. Rau, and M. Biehl. The Statistical Mechanics of Learning a Rule. *Rev. Mod. Phys.*, 65:499–556, 1993.

[163] N. M. Weinberger. The nucleus basalis and memory codes: Auditory cortical plasticity and the induction of specific, associative behavioral memory. *Neurobiology of Learning and Memory*, 80(3):268 – 284, 2003. Acetylcholine: Cognitive and Brain Functions.

[164] Y. M. Wen and B. L. Lu. Incremental learning of support vector machines by classifier combining. In *Proc. of 11th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2007)*, volume 4426 of *LNCS*, 2007.

[165] G. Wu, W. Xu, and H. Leng. Inexact and incremental bilinear Lanczos components algorithms for high dimensionality reduction and image reconstruction. *Pattern Recognition*, 48(1):244–263, 2015.

[166] X. Wu, P. Rózycki, and B. Wilamowski. A hybrid constructive algorithm for single-layer feedforward networks learning. *IEEE Transactions on Neural Networks and Learning Systems*, 26(8):1659–1668, 2015.

[167] W. Xi-Zhao, S. Qing-Yan, M. Qing, and Z. Jun-Hai. Architecture selection for networks trained with extreme learning machine using localized generalization error model. *Neurocomputing*, 102:3–9, 2013.

[168] J. Xin, Z. Wang, L. Qu, and G. Wang. Elastic extreme learning machine for big data classification. *Neurocomputing*, (Part A):464–471, 2015.

[169] Y. Xing, F. Shen, C. Luo, and J. Zhao. L3-SVM: A lifelong learning method for SVM. In *IJCNN*, volume 2015-September, 2015.

[170] H. Yang, S. Fong, G. Sun, and R. Wong. A very fast decision tree algorithm for real-time data mining of imperfect data streams in a distributed wireless sensor network. *International Journal of Distributed Sensor Networks*, 2012, 2012.

[171] G. Yin, Y.-T. Zhang, Z.-N. Li, G.-Q. Ren, and H.-B. Fan. Online fault diagnosis method based on incremental support vector data description and extreme learning machine with incremental output structure. *Neurocomputing*, 128:224–231, 2014.

[172] X.-C. Yin, K. Huang, and H.-W. Hao. DE2: Dynamic ensemble of ensembles for learning nonstationary data. *Neurocomputing*, 165:14–22, 2015.

[173] X.-Q. Zeng and G.-Z. Li. Incremental partial least squares analysis of big streaming data. *Pattern Recognition*, 47(11):3726–3735, 2014.

[174] C. Zhang, R. Liu, T. Qiu, and Z. Su. Robust visual tracking via incremental low-rank features learning. *Neurocomputing*, 131:237–247, 2014.

[175] H. Zhang, P. Wu, A. Beck, Z. Zhang, and X. Gao. Adaptive incremental learning of image semantics with application to social robot. *Neurocomputing*, 173:93–101, 2016.

[176] H. Zhang, X. Xiao, and O. Hasegawa. A load-balancing self-organizing incremental neural network. *IEEE Transactions on Neural Networks and Learning Systems*, 25(6):1096–1105, 2014.

[177] R. Zhang, Y. Lan, G.-B. Huang, and Z.-B. Xu. Universal approximation of extreme learning machine with adaptive growth of hidden nodes. *IEEE Transactions on Neural Networks and Learning Systems*, 23(2):365–371, 2012.

[178] X. Zhou, Z. Liu, and C. Zhu. Online regularized and kernelized extreme learning machines with forgetting mechanism. *Mathematical Problems in Engineering*, 2014, 2014.

[179] M. Zuniga, F. Bremond, and M. Thonnat. Hierarchical and incremental event learning approach based on concept formation models. *Neurocomputing*, 100:3–18, 2013.