Uncertainty-based Deep Learning Networks for Limited Data Wetland User Models

Andrew Hoblitzell

December 12, 2018

C13: Uncertainty-based Deep Learning Networks for Limited Data Wetland User Models

***Abstract*—A method for dealing with limited data in deep networks is given which is based upon calculating uncertainty associated with remaining data. The method was developed for the Watershed REstoration using Spatio-Temporal Optimization of REsources (WRESTORE) system, an interactive decision support system designed for performing multi-criteria decision analysis with a distributed system of conservation practices on the Eagle Creek Watershed. Eventually, these results from neural network user models may be integrated in to an existing genetic algoritm to find Pareto optimal solutions for multiple stakeholders within the constraints of the physical and socio-economic environment. Our results show faster and more stable convergence when using an uncertainty based incremental sampling method than when using a standard random incremental sampling method. This work describes the existing WRESTORE system, provides details about the implementation of our uncertainty based incremental sampling method, and provides discussion of our results and future work.**

***Index Terms*—multiple criteria decision analysis; limited data; data science; active learning, optimal design, incremental learning, deep learning, interactive machine learning, design of experiments, optimal design, online learning, optimal decisions, systems engineering, statistical process control, adaptive learning, iterative learning, incremental learning, uncertainty quantification, simulation and modeling, statistical efficiency, Tensorflow, massively parallel learning, graphical processing unit**

I. Introduction

WRESTORE is a user-friendly and interactive web-based decision support system, which is used to help the landowners and other stakeholders including the government to help them in the implementation of conservation practices. By using the system, users can identify substitutions according to their needs. Eagle Creek Watershed is a flat-land covering about 162 square miles located in Central Indiana, which is almost 10 miles northwest of downtown Indianapolis. Eagle Creed Watershed can be divided into sub-basins. There are approximately 130 sub-basins connected to one another based on the geography illustrated in Geological Survey topography map and are used to measure the flow of water in the district.

Wetlands are either wet throughout the year, during certain climatic spells or during daytime. These wetlands are used to support the crops and the vegetation in the area by keeping the soil saturated. Wetlands have many benefits, such as reducing the overall threat of flooding, by reducing peak water flow. Wetlands also provide other benefits, such as absorbing undesirable inorganic materials like fertilizers. [1] The United States Department of Agriculture has created several programs for impriving wetlands such as the Wetlands Reserve Program. [2]

Other best management practices for wetlands include filter strips, grassed waterways, crop rotation, no-till, strip cropping and cover crops. Filter strips remove residue, organic materials, and various other pollutants from wastewater and runoff water and provide benefits such as removing residue before it enters the water way. [3] Grassed waterways carry runoff from rigorous flow and thus effectively help with soil erosion. [4] Crop rotation usually consists of growing of food crops mostly with various field crops and offers benefits like reduction of runoff. [3] No-till ceases planned man-made agitations like digging, stirring, and overturning and has environmental advantages while also reducing the labor costs of human intervention on the land. [5] Strip cropping is generally used on natural or recreational lands and provides benefits like increased soil moisture and water quality. [6] Cover crops, which include ryegrass and oats, are grown between regular crops and improve the overall productivity of farmers crops. [7]

Interactive machine learning exists when an algorithm to queries a user, an environment, or some other data source to obtain feedback relative to its performance on an objective function. In situations where a user is involved, it often becomes expensive to obtain a large amount of feedback due to issues like user fatigue, a large amount of research is often thus focused on finding the best methods for optimally querying domain experts. Interactive machine learning goes by many names in literature, including active learning, online learning, incremental machine learning, design of experiments, and and optimal design.

There are many approaches which exist in the literature for discovering optimal points for user query. One method, model based uncertainty, involves determining which points the implemented model is the most uncertain about. Another method, based around ensemble learning, involves determining where the most uncertainty exists within a collection of models. Other models involve random exploration, parameterized exploration versus exploitation, or methods which minize prediction stability, generalization error, etc. Such methods require less queries of the user, and are generally very highly model dependent, thus it is also desirable to empirically measure improvement with any such developed methods.

II. Background Literature

There has been increased interest from domain experts in several fields in algorithms which facilitate machine learning coupled with human interaction. It has become apparent that machines can interact with people to solve problems more efficiently than autonomous systems would be able to do in solitude. Applying a human-centered perspective to machine learning and exploring the co-adaptation of humans and machine learning algorithms leads to much more effective outcomes. This section will briefly explore ways that other researchers in the field have performed interactive and incremental machine learning.

Luo et al. used empirical eigenfunctions and neural networks to approximate optimal control with parabolic partial differential equations. [13] Their contributon makes use of a neural network to approximate the nonlinear cost function of an HJB-like equation. Luo et al. demonstrate and prove the convergence of their neural network based approach, although they do not fully show the well-posedness of the partial differential equation system nor the impact of neural network approximation error on the controller design.

Kim et al. developed a cerebellar model arithmetic computer neural network for optimal control. [14] Kim et al.'s approach copes with nonlinearities and provides explicit solutions to the Hamilton-Jacobi-Bellman (H-J-B) equation. While the approach is somewhat limited by approximation, the use of cerebellar model arithmetic computer neural network helps with reducing the typical amount of necessary parametrization.

Tsai et al. propose a method for optimizing fuzzy neural networks which determines the optimal learning rate to minimize the next-step mean error. [15] Tsai et al. approach reduces the effect of outliers to more statistically efficiently perform function approximation, although their experimental results are not fully depicted due to the space allotted to them. Tsai et al. provide an example of how to more efficiently perform the prediction task.

Choi et al. compare artificial neural networks, grey models, Markov regime switching, and other methods for the purpose of color trend forecasting within the fashion industry. [16] Choi's approach for dealing with limited data offers the advantage of determining model statistical efficiency as well as model speed and other characteristics; however, it does not offer the advantage of improving an individual model's efficiency. It is important to note that Choi's analysis shows that artificial neural networks can be effective with the presence of limited data.

Cauwenberghs et al. describe a method for incremental learning and decremental unlearning methods for training a SVM. [17] The decremental learning approach uses a leave one out method. The article establishes that the decremental unlearning and incremental learning offers efficient scheme for perfect SVM training online and leave one out considerations, however the article could be strengthened with more detailed proofs and simulation results. In addition, geometric interpretation of decremental unlearning and more complex data sets could be studied to further support the validity of their approach.

Carpenter presents a Fuzzy ARTMAP architecture for analysis of neural networks. [18] The presented architecture utilizes fuzzy logic and Adaptive Resonance Theory (ART) and develops a new minimax learning rule. Several training attempts are performed within the paper in order for the algorithm to minimize predictive error and improve accuracy. The algorithm encounters some issues with letter image data and also runs in to issues with some voting criteria, but in general the approach provides significant and stable accuracy.

Langkvist et al. applied interactive imaging to convolutional neural networks when working on image labeling tasks. [19] One very common problem with image classification problems is the absence of significant levels of satisfactory amounts of labeled data. To deal with the problem of unnecessary amounts of label data, a human-in-the-loop intelligent system which allows the algorithm and user to collaborate is developed. Although the approach shows limited results with human face imaces, in general the approach works well with some other data sets. Future work identified by the authors include GPU optimizations and dynamically adjusting the size of the convolutional neural network.

Amershi et al. demonstrate the importance of studying the users of interactive machine learning systems. [20] Several studies which demonstrate how colearning results in a tight coupling between the algorithm and the user, how some systems fail to account for colearning, and identifying new potential avenues for human algorithm example. The authors eschew the typical iterative tuning of parameters and note that interactive approaches allow the user to direct the algorithm in more rapid, focused, and incremental ways. The authors note that users value transparency and understandability in learning systems, which can even help lead to better performance again objective criterion. The authors conclude that there are still many challenges present in the field of interactive machine learning.

III. WRESTORE AND INTERACTIVE OPTIMIZATION

As briefly discussioned in the introduction, the WRESTORE system is an interactive decision support system designed for balancing the interests of multiple stakeholders with respect to several quantitative and qualitative criteria relevant to the watershed design problem. Many strictly quantitative approaches do not fully consider qualitative information that affect suitability of different watershed designs. The Interactive Genetic Algorithm with Mixed Initiative Interaction (IGAMII) is part of the WRESTORE system which more fully takes this in to account. Due to space limitations, it is not possible to fully consider the IGAMII algorithm implementation, WRESTORE system design, or decision maker (DM) approach; however, all of this is more fully described in the cited papers.[8] [9] [10]

Previous work in the system has included the training of virtual decision makers or user models which will offer the perspective of different stakeholders involved in the interactive optimization process. Several different approachers for user modeling were implemented and compared, including traditional artificial neural networks, fuzzy logic, and deep neural networks. Comparison of several different implementations and configurations led to the conclusion that a deep neural network approach offered the most accuracy for virtual decision making; however, it also quickly became apparent that one of the chief challenges with such an approach is the lack of training data which is available. More detail about this work is available in the cited paper.[11]

The implementation of this system has seen practical usage for outcomes on actual stakeholders. The system is implemented on the Eaglecreek Creek Watershed (ECW) 10 miles northwest of Indianapolis, Indiana, in the United States of America. Indiana lost more than 85% of its wetland area due to urbanization by the 1980s, leading to the need to recreate areas where downstream flood water can be stored. Several approaches are considered as potential ecological solutions for increasing the capacity of watersheds to store runoff waters upstream, and decrease overall flooding downstream. The outcomes of the WRESTORE system have been used to propose designs which would offer effective wetland design parameters which are spatially optimized to use fewer sites, smaller wetlands, and reduced financial costs. More details about the benefits offered and Pareto optimization against different constraints are available in the cited paper.[12]

IV. Methodology

In this section, we further describe the approach we have implemented for deep learning[21] and the different presented sampling techniques.

1) Design of the Neural Network: Our implementation makes use of a unified dataflow graph to represent both computation and state. Nodes represent operations which need to be performed, while edges represent data flowing along a graph. The advantages of this approach include massive parallelization and lazy evaluation.[22] TensorFlow is the Apache-licensed reference implementation commonly used for deep learning which we utilized. Scripts written in the framework can be executed in environments as heterogeneous as a single phone to thousands of computational devices such as GPU cards. The library has been used across in speech recognition, computer vision, robotics, information retrieval, natural language processing, geographic information extraction, and computational drug discovery, and other areas.

Our input layer consisted of environmental fitness functions. The output layer was the prediction of the user model given the input environmental fitness functions.

Activation state of each unit within each layer, X_i, was given by a value of 0 for not activated and a value of 1 for activated. The weight given between unit i and unit j is given by weight W_{ij}. The activation of a unit is given by the sum of the products of output Y_j and weight W_{ij} along with the bias term, b_j:

$$X_i = \sum_j W_{ij} Y_j + b_j.$$

Given activation X_i, output Y_i is computed using the activation function. Output Y_i is propagated in a feedforward fashion to the next layer until it reaches the output layer. This process is depicted in Fig. 1.

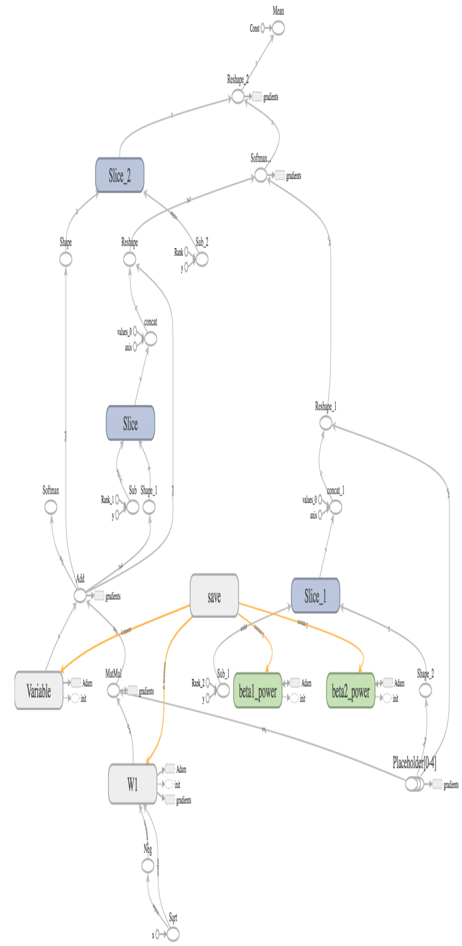

Fig. 1. Visualization of implemented NN architecture

Our implementation of stochastic gradient descent made use of the first-order gradient-based Adam optimizer.[23] Adam is based on estimates of lower-order moments, and is invariant to rescaling of the gradients which makes it computationally efficient, lightweight in memory, and ideal for large parameter optimization. Adam may also be used for non-stationary objectives and problems with noisy gradients, which is ideal for our subjective objective functions. Stochastic gradient descent was performed on the cross-entropy function:

$$\hat{R} = -\sum_{i=1}^{n} \sum_{g=1}^{G} Y_{ig} \log(\hat{\pi}_g(X_i))$$

We chose the rectified linear unit as our activation function because of its resemblance to the biological domain and its current prevalence within the literature. Advantages of linear rectifiers include their sparse activation, the fact that being dominantly addition and multiplication make them fast on graphical processing units, and the fact that they are efficient at gradient propagation.

The softmax function, or normalized exponential function, is the gradient-log-normalizer of the categorical probability

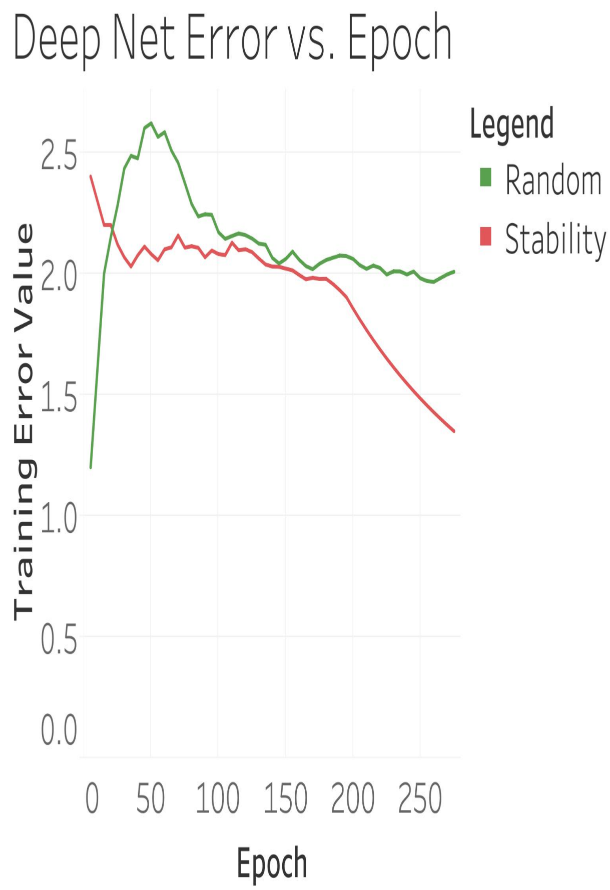

Fig. 2. Training Error vs. Epoch

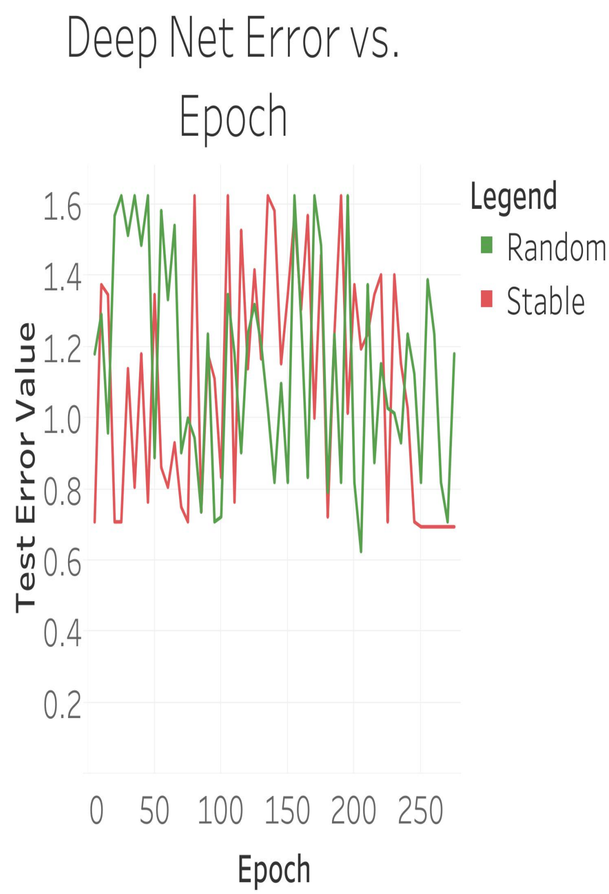

Fig. 3. Test Error vs. Epoch

distribution and is often used in multiclass classification methods and multinomial logistic regression. We use the softmax layer as our final layer because the softmax layer gives us a list of values between 0 and 1 that add up to 1, which makes it trivial to determine the prediction of the neural network in an intuitive way.

Zheng et al. developed a general stability training method to stabilize deep networks against extreme instability against contrived input perturbations commonly known as adversarial examples.[24] Zheng et al. develop a parameterized Gaussian weighted stability noise factor which we replicate the logic of here for comparison against a new uncertainty based sampling method. The uncertainty based sampling method proposed works by taking the most uncertain training examples from the softmax function within each mini-batch. The next section will compare the results of a random sampling method, a stability based sampling method, and our uncertainty based sampling method.

V. Results

A. Deep Learning

Our deep learning approach used a multi-layer neural network. Our implementation performed a hybrid grid search and random search over the regularization factor, dropout ratio, the size of hidden layers, and number of epochs.

There are many activator functions in the community, and each one has its own benefits and drawbacks. We elected to use rectified linear units which are formally described as f(i)=max(o,i).

Overfitting is a nontrivial problem in multi-layer neural networks, especially implementations with constraints on the amount of data which they have access to. Several techniques exist for dealing with limited data including regularization and randomly dropping connections and their units from the neural network to prevent high coupling between nodes. We utilized both strategies in our training.

Autoencoders are a commonly used technique for learning dimensionality reduced versions of a search space; we did not need to utilize this technique since our problem domain did not have a feature space which would currently require such techniques.

B. Summary of Results

We used the mean absolute error (MAE) as the error performance measure for this analysis, which is given by the

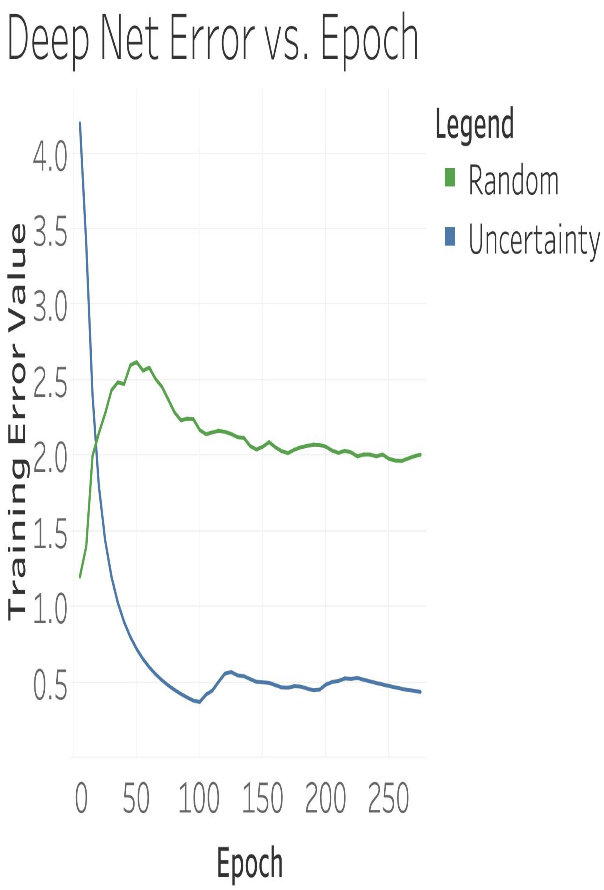

Fig. 4. Training Error vs. Epoch

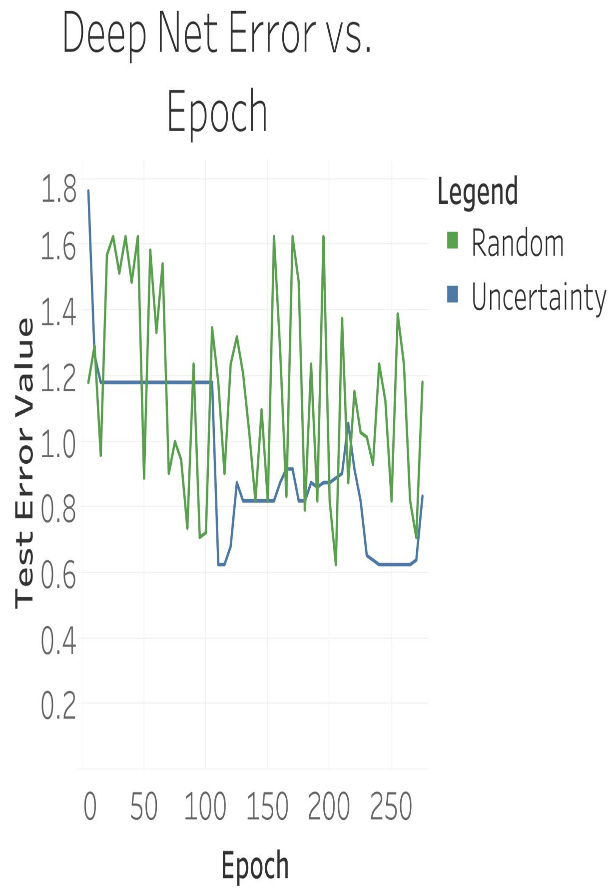

Fig. 5. Test Error vs. Epoch

following :

$$\frac{1}{N}\sum_{i=1}^{N}|T_i - C_i|$$

where N is the number of samples in the test set, T_i is the true label of the i-th sample and C_i is the predicted label.

Mean absolute error has the advantage of offering the average absolute difference between C_i and T_i, which is advantageous to others in the community because its interpretation is clearer than similar measures like root mean squared error. MAE has the disadvantage of being a scale-dependent accuracy measure which cannot be used to make comparisons between series of different scales; however, all of the comparisons made here are within the same scale.

As can be seen in Figures 2 and 3, the training error and test errors for stability based incremental sampling outperformed random based sampling in both cases. In addition, as can be seen in Figures 4 and 5, the training error and test errors for uncertainty based incremental sampling also outperformed random based sampling in both cases. In both cases, the error rates of non-random based sampling provided generally faster convergence as well. The experiments were ran with microbatch size = 1 and learning rate = 0.008.

In the most interesting case, we can compare the training and test errors of uncertainty based incremental sampling and stability based incremental sampling. The proposed uncertainty based sampling method was able to converge with half of the training data of the stability based approach. The uncertainty based approach and stability showed relatively similar values for overall test error, which we would expect when they have ran over the entire training corpus.

C. Future Work

There are many areas where we believe research in this area can go next. Random exploration, parameterized exploration versus exploitation, methods which minimize prediction stability, methods which minimize prediction generalization error, and other techniques are all design of experiment based techniques which could be explored further when it comes to sampling technique. In addition, there are also other techniques which could be compared against which exist within current deep learning literature.

From a more practical perspective, we also place a great emphasis on embedding these techniques back in to our wetland design interactive decision support system. Implementing such techniques within our system would allow for less expensive querying of the user and faster convergence to the Pareto-optimal solutions most closely aligned with our users' preferences, which can ultimately drive better community and stakeholder outcomes.

VI. Conclusion

In this paper, we developed an uncertainty based sampling technique for the Watershed REstoration using Spatio-Temporal Optimization of REsources (WRESTORE) system to more efficiently deal with the expensive task of user querying. The proposed uncertainty based sampling method showed faster convergence than both a random sampling technique and a stability based sampling technique. There are several areas remaining for future work, including comparison against other sampling based techniques and integration back in to our interactive decision support system.

Acknowledgment

This work received support from the National Oceanic and Atmospheric Administration as well as National Science Foundation award ID #1332385 (previously ID# 1014693). We would like to acknowledge and thank Ms. Jill Hoffmann for her help with organizing multiple workshops with stakeholders to enable testing of the tool. In addition, we deeply appreciate the effort of the reviewers whose feedback greatly improved the quality of our manuscript.

References

[1] "Wetlands — US EPA", US EPA, 2018. [Online]. Available: https://www.epa.gov/wetlands.

[2] "United States Department of Agriculture Natural Resources Conservation Service: NRCS — NRCS", In.nrcs.usda.gov, 2018. [Online]. Available: http://www.in.nrcs.usda.gov/programs/WRP/WRPhomepage.html.

[3] "United States Department of Agriculture Natural Resources Conservation Service: ", Nrcs.usda.gov, 2018. [Online]. Available: https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_006812.pdf.

[4] "United States Department of Agriculture Natural Resources Conservation Service: Report on CRP Practice CP8A", Nrcs.usda.gov, 2018. [Online]. Available http://www.in.nrcs.usda.gov/programs/CRP/2012_job_sheet/IN%20CP8A%20Program%20Sheet%20vMar1-0.pdf

[5] "Conservation Practices that Save: Crop Residue Management", Nrcs.usda.gov, 2018. [Online]. Available http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs143_023624.pdf

[6] "United States Department of Agriculture: Conservation Practice Overview", Nrcs.usda.gov, 2018. [Online]. Available http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs143_026169.pdf

[7] "Natural Resources Conservation Service Conservation Practice Cover Crop", Nrcs.usda.gov, 2018. [Online]. Available http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1046845.pdf

[8] M. Babbar-Sebens, B.S. Minsker, "Interactive genetic algorithm with mixed initiative interaction for multi-criteria ground water monitoring design", Applied Soft Computing, vol. 12, no. 1, pp. 182-195, 2012.

[9] Babbar-Sebens, M., Mukhopadhyay, S., Singh, V.B. and Piemonti, A.D., 2015. A web-based software tool for participatory optimization of conservation practices in watersheds. Environmental Modelling & Software, 69, pp.111-127.

[10] Piemonti, A.D., BabbarSebens, M., Mukhopadhyay, S. and Kleinberg, A., 2017. Interactive genetic algorithm for usercentered design of distributed conservation practices in a watershed: An examination of user preferences in objective space and user behavior. Water Resources Research, 53(5), pp.4303-4326.

[11] Andrew Hoblitzell, Meghna Babbar-Sebens, Snehasis Mukhopadhyay, "Fuzzy and deep learning approaches for user modeling in wetland design", IEEE International Conference on Systems Man and Cybernetics(SMC), 2016.

[12] BabbarSebens M, Barr RC, Tedesco LP, Anderson M. 2013. Spatial identification and optimization of upland wetlands in agricultural watersheds. Ecological Engineering 52: 130142.

[13] B. Luo and H. N. Wu, "Approximate Optimal Control Design for Nonlinear One-Dimensional Parabolic PDE Systems Using Empirical Eigenfunctions and Neural Network," in IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 42, no. 6, pp. 1538-1549, Dec. 2012. doi: 10.1109/TSMCB.2012.2194781 http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6198365&isnumber=6353142

[14] Y. H. Kim and F. L. Lewis, "Optimal design of CMAC neural-network controller for robot manipulators," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 30, no. 1, pp. 22-31, Feb 2000. doi: 10.1109/5326.827451 http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=827451&isnumber=17956

[15] Hung-Hsu Tsai and Pao-Ta Yu, "On the optimal design of fuzzy neural networks with robust learning for function approximation," in IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 30, no. 1, pp. 217-223, Feb 2000. doi: 10.1109/3477.826964 http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=826964&isnumber=17934

[16] T. M. Choi, C. L. Hui, S. F. Ng and Y. Yu, "Color Trend Forecasting of Fashionable Products with Very Few Historical Data," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 6, pp. 1003-1010, Nov. 2012. doi: 10.1109/TSMCC.2011.2176725

[17] G. Cauwenberghs and T. Poggio, "Incremental and Decremental Support Vector Machine Learning", Advances in Neural Information Processing Systems 13 (NIPS 2000), pp. 388-394, 2000.

[18] G. Carpenter, "Fuzzy ARTMAP: A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps", Cognitive & Neural Systems CAS/CNS Technical Reports, 1991.

[19] M. Langkvist, M. Alirezaie, A. Kiselev, and A. Loutfi. Interactive learning with convolutional neural networks for image labeling. In International Joint Conference on Artificial Intelligence (IJCAI), New York, USA, 9-15th July, 2016, 2016.

[20] Amershi, S., Cakmak, M., Knox, W.B., Kulesza, T.: Power to the people: the role of humans in interactive machine learning. AI Mag. 35(4), 105120 (2014).

[21] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436.

[22] M. Abadi et al., "TensorFlow: A System for Large-Scale Machine Learning TensorFlow: A system for large-scale machine learning", 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI'16), pp. 265-284, 2016.

[23] Kingma, Diederik P. and Ba, Jimmy. Adam: A Method for Stochastic Optimization. arXiv: 1412.6980 [cs.LG], December 2014.

[24] Zheng, S., Song, Y., Leung, T., Goodfellow, I.: Improving the robustness of deep neural networks via stability training. In: CVPR 2016 (2016)