

# Abstract

Normal neural networks trained with gradient descent and back-propagation have received great success in various applications. On one hand, point estimation of the network weights is prone to over-fitting problems and lacks important uncertainty information associated with the estimation. On the other hand, exact Bayesian neural network methods are intractable and non-applicable for real-world applications. To date, approximate methods have been actively under development for Bayesian neural networks, including but not limited to, stochastic variational methods, Monte Carlo dropouts, and expectation propagation. Though these methods are applicable for current large networks, there are limits of these approaches with either under estimation or over-estimation of uncertainty. Extended Kalman filters (EKFs) and unscented Kalman filters (UKFs), which are widely used in data assimilation community, adopt a different perspective of inferring the parameters. Nevertheless, EKFs are incapable of dealing with highly non-linearity, while UKFs are inapplicable for large network architectures.

Ensemble Kalman filters (EnKFs) serve as great methodology in atmosphere and oceanology disciplines targeting extremely high-dimensional, non-Gaussian, and nonlinear state-space models. So far, there is little work that applies EnKFs to estimate the parameters of deep neural networks. By considering neural network as a nonlinear function, we augment the network prediction with parameters as new states and adapt the state-space model to update the parameters. In the first work, we describe the ensemble Kalman filter, two proposed algorithms for training both fully-connected and Long Short-term Memory (LSTM) networks, and experiment it with a synthetic dataset, 10 UCI datasets, and a natural language dataset for different regression tasks. To further evaluate the effectiveness of the proposed training scheme, we trained a deep LSTM network with the proposed algorithm, and applied it on five real-world sub-event detection tasks. With a formalization of the sub-event detection task, we develop an outlier detection framework and take advantage of the Bayesian Long Short-term Memory (LSTM) network to capture the important and interesting moments within an event. In the last work, we develop a framework for student knowledge estimation using Bayesian network. By constructing student models with Bayesian

network, we can infer the new state of knowledge on each concept given a student. With a novel parameter estimate algorithm, the model can also indicate misconception on each question. Furthermore, we develop a predictive validation metric with expected data likelihood of the student model to evaluate the design of questions.