Chuong Ngo - Summary for Eric Eaton’s Colloquium

Online Multi-Task Learning via Sparse Dictionary Optimization

**Summary**

Traditionally, machine learning structures are typically trained independent of other, possibly related, tasks. This can prove to be very inefficient since the models are trained one at a time. Multi-task Learning (MTL) is a different approach to training machine learning models that tries to correct some of the inefficiencies alluded to prior. MTL trains multiple models in parallel on the same training set with shared representation. This can lead a better trained model for the primary tasks due to commonality and overlaps between the various tasks. It also helps with regularization, allowing the models to work better with ill-formed input and helping to prevent overfitting. However, MTL is very computationally expensive and is more suitable for pre-training/offline training as opposed to online training.

A popular structure for machine learning is the support vector machine (SVM). SVMs are supervised learning models that, when presented with input data, classify the data as belonging to either one of two categories. Ideally, this will result in model that, when represented as a set of points in space, will show its data as clearly defined with as large of a distance between the two categories as is reasonable. This is further helped by mapping the SVMs to higher dimensional spaces, increasing the sparsity of data space and helping to achieve a strong differentiation between the two categories. This is done using the “kernel trick”, which allows for the computation of high dimensional, implicit features without computing the data coordinates in that space. This decreases the computational load. That is the key.

In order to do be effective as an online machine learning model, the algorithm used to create the model must be able to take input in as training data as it comes in. Therefore, those algorithms must be efficient. This is why conventional MTL is not a good candidate for online models. By using SVMs for MTL, as opposed to other more complex structures, the computational complexity of the model is reduced, making for an algorithm that is more suited for online learning. In order to adopt SVMs that are normally used to learn a dictionary for sparse coding of input data points to one that can be used to learn a dictionary for sparse coding of parameter vectors, a generalized singular value decomposition that accepts two symmetric positive semidefinite matrix in different data spaces. Adapting the new MTL-SVD further to lifelong learning yields ELLA-SVD, which has high performance and good accuracy on a number of tasks, including landmine detection. However, other tasks, like facial recognition, showed that a modified ELLA-SVD, called ELLA Incremental, was preferable. ELLA Incremental updates the various columns of the matrix independently when new data is received. This introduces additional overhead to the algorithm, but has higher accuracy for some tasks when compared to ELLA-SVD. The author’s hybrid approach of combining both ELLA-SVD and ELLA Incremental proved to be the most successful.

**Reference**

Ruvolo, P., & Eaton, E. (2014, June). Online Multi-Task Learning via Sparse Dictionary Optimization. In *Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI-14)*.