The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)

An SVM-Based Framework for Long-Term Learning Systems

Diana Benavides-Prado
Department of Computer Science
The University of Auckland
dben652@aucklanduni.ac.nz

Abstract

In our research, we study the problem of learning a sequence of supervised tasks. This is a long-standing challenge in machine learning. Our work relies on transfer of knowledge between hypotheses learned with Support Vector Machines. Transfer occurs in two directions: forward and backward. We have proposed to selectively transfer forward support vector coefficients from previous hypotheses as upper-bounds on support vector coefficients to be learned on a target task. We also proposed a novel method for refining existing hypotheses by transferring backward knowledge from a target hypothesis learned recently. We have improved this method through a hypothesis refinement approach that refines whilst encouraging retention of knowledge. Our contribution is represented in a long-term learning framework for binary classification tasks received sequentially one at a time.

1 Introduction

Learning a sequence of tasks is a long-standing challenge in machine learning. Paradigms such as learning to learn, early lifelong learning and metalearning have acknowledged this problem. Recently, Chen and Liu (2016) formalised lifelong machine learning as a process composed of a set of related tasks that arrive sequentially and share knowledge. They identified three core characteristics of these systems: 1) to learn new tasks better, supported by existing knowledge; 2) to store knowledge continuously and incrementally in a knowledge base; 3) to perform continuous learning. A variety of research in transfer, hypothesis transfer, multitask, meta and deep learning has explored the first characteristic. Research in lifelong learning and related areas have studied the second characteristic. The last property, that should ideally pursue refinement of existing knowledge, has only been explored recently (Ruvolo and Eaton 2013; Fei, Wang, and Liu 2016). Continual learning with deep neural networks has focused on the challenge of learning new tasks without forgetting existing knowledge (Yoon, Yang, and others 2017). Nevertheless, the problem of refining existing knowledge has been scarcely explored.

In our research we study a framework for lifelong machine learning with these three properties. Our framework is

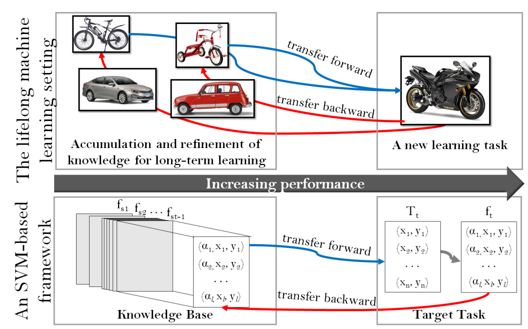

Figure 1: The lifelong machine learning problem (top) and a framework based on SVM (bottom).

built on top of Support Vector Machines (SVM) for classification tasks and some of its variants. The first characteristic is tackled by a method that transfers selected knowledge from a set of source hypotheses to a target task. Experimentally we demonstrated that our method speeds the convergence rate of the target task. A second method proposed to *transfer backward*, a novel ability of lifelong learning systems that aims to refine hypotheses from previous tasks. These hypotheses are stored in a knowledge base. Experiments with small real-world datasets denoted a potential of this approach for continuously improving performance of existing hypotheses. We have extended this method to systems composed of any number of tasks. This novel approach encourages retention of knowledge while refining existing hypotheses. Experiments with large synthetic and real-world datasets demonstrated the feasibility of this approach. Figure 1 sketches a framework based on selective transfer forward and backward. A brief explanation of the two main components of this framework is provided in Sections 2 and 3.

2 Selective Transfer Forward

For transferring forward, we have proposed to use elements of existing hypotheses $f_s \in S$ to aid learning of a target task T_t. Support vectors from previous SVM hypotheses are selected, and their coefficients aggregated and used to upper-bound coefficients of support vectors to be learned as part of a target hypothesis f_t on T_t. This problem has been formalised as an SVM classification problem with a modified constraint (Benavides-Prado, Koh, and Riddle 2017):

$$\max_{\alpha} F(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$
$$s.t. \sum_{i=1}^{n} y_i \alpha_i = 0, \forall i \; 0 \leq \alpha_i \leq C + c_i, c_i = \frac{|F|}{|S|} \sum_{k=1}^{s} \alpha_k \tag{1}$$

Here, a coefficient α_i for f_t is upper-bounded by the constraint $C + c_i$, which is composed of the original upper-bound C and c_i, an aggregation of $\alpha = \{\alpha_1, \ldots, \alpha_s\}$ coefficients transferred from source support vectors $x_s \in F$ that are similar to a target x_i training example. F is the subset of existing f_s hypotheses related to the task T_t, and s is the number of support vectors in that set that are similar to x_i. As a result, training examples on the target task that are more closely resembled by support vectors in previous tasks contribute more to the optimization problem.

3 Selective Transfer Backward

The problem of transferring backward aims to refine existing SVM hypotheses f_s by exploiting knowledge collected while transferring forward to learn a target f_t. From the method in Section 2, tuples of the form:

$$< (x_s, y_s, \alpha_s), (x_t, y_t, \alpha_t) > \tag{2}$$

can be conformed. Here, (x_s, y_s, α_s) corresponds to a support vector from a source hypothesis f_s, and (x_t, y_t, α_t) corresponds to a support vector from the target hypothesis f_t learned recently, which were involved in transfer in Eq. 1. These tuples represent subspaces of shared knowledge between f_s and f_t, and can be potentially used for refining the existing f_s. We propose to approach this refinement by solving a modified SVM classification problem as follows:

$$\max_{\alpha} F(\alpha) = -\frac{1}{2}[(1-\Gamma) \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$
$$+\Gamma \sum_{i,k=1}^{l,2o} \alpha_i y_i \alpha_k^o K(x_k^o, x_i)] \tag{3}$$
$$s.t. \sum_{i=1}^{l} y_i \alpha_i = 0, \sum_{i=1}^{l} \alpha_i \geq \nu, \forall i \; 0 \leq \alpha_i \leq 1/l$$

which is based on ν-SVM (Schölkopf et al. 2000). ν-SVM for classification tasks is an alternative that considers a parameter ν that limits both the degree of compression of an SVM hypothesis, acting as a lower bound on the fraction of support vectors, and the training error, acting as an upper bound on the fraction of margin errors. In our method, refinement is controlled by controlling the training error, whilst retention of knowledge is controlled by controlling the compression. The term $\sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$ describes the space of the current $x_s \in f_s$. The term $\sum_{i,k=1}^{l,2o} \alpha_i y_i \alpha_k^o K(x_k^o, x_i)$ describes the space of the f_s source hypothesis space intersected with the f_t target hypothesis space. Here, α_k^o and x_k^o, with $1 \leq k \leq 2o$, are extracted from o functions f_o learned with one-class SVM. Each of these functions uses as training examples the elements of a tuple represented as in Eq. 2. The parameter Γ, set generally small, controls the contribution of the last term.

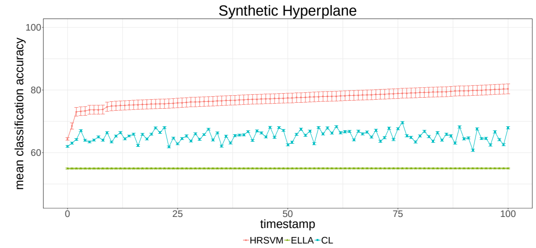

Figure 2: Mean classification accuracy at each timestamp. Error bars show 95% confidence intervals. t_0 is the test performance after half of the tasks have been learned.

4 Experimental Results

Our framework has been tested in synthetic and real-world data. Figure 2 shows example results for learning systems that learn hyperplanes. We evaluated our method (HRSVM) and counterparts (ELLA and CL). In this example, synthetic training and test sets for several hyperplanes were generated using an existing method[1]. At each timestamp, a new hyperplane task is learned, existing hypotheses are refined and the knowledge base is updated. A learning system should denote better performance as the sequence of tasks progresses.

5 Future Work

Some avenues of research derived from our work are: 1) extension to multi-class settings for groups of tasks learned sequentially, 2) continual learning methods that pursue refinement, 3) studying performance metrics for long-term learning systems, 4) investigating the impact of aspects such as relatedness of tasks, number and quality of tasks.

6 Acknowledgements

Special thanks to my supervisors, Dr Patricia Riddle and Dr Yun Sing Koh, at the Department of Computer Science, The University of Auckland, New Zealand.

References

Benavides-Prado, D.; Koh, Y. S.; and Riddle, P. 2017. AccGenSVM: Selectively Transferring From Previous Hypotheses. In *IJCAI*, 1440–1446. AAAI Press.

Chen, Z., and Liu, B. 2016. Lifelong Machine Learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 10(3):1–145.

Fei, G.; Wang, S.; and Liu, B. 2016. Learning Cumulatively to Become More Knowledgeable. In *KDD*, 1565–1574.

Ruvolo, P., and Eaton, E. 2013. ELLA: An Efficient Lifelong Learning Algorithm. *ICML* 28:507–515.

Schölkopf, B.; Smola, A. J.; Williamson, R. C.; and Bartlett, P. L. 2000. New Support Vector Algorithms. *Neural Computation* 12(5):1207–1245.

Yoon, J.; Yang, E.; et al. 2017. Lifelong Learning with Dynamically Expandable Networks. *arXiv:1708.01547*.

[1]http://scikit-learn.org/stable/auto_examples/svm/plot_separating_hyperplane.html