A GRU-based Meta-learning Model Based on Active Learning

Honglan Huang
College of Systems Engineering, National University of Defense Technology
Changsha, China
e-mail: huanghonglan17@nudt.edu.cn

Shixuan Liu
College of Systems Engineering, National University of Defense Technology
Changsha, China
e-mail: liushixuan19@nudt.edu.cn

Yanghe Feng
College of Systems Engineering, National University of Defense Technology
Changsha, China
e-mail: fengyanghe@yeah.net

Jincai Huang
College of Systems Engineering, National University of Defense Technology
Changsha, China
e-mail: huangjincai@nudt.edu.cn

Zhong Liu
College of Systems Engineering, National University of Defense Technology
Changsha, China
e-mail: liuzhong@nudt.edu.cn

***Abstract*—In the realities of machine learning, labeling a data set may be expensive, tedious, or extremely difficult and it is often not easy to choose the common criteria for active learning to select samples for different data sets. In order to solve these difficulties, this paper introduces a Gated Recurrent Unit (GRU)-based meta-learner model, which combines active learning with reinforcement learning and uses it in a stream-based one-shot learning task. Based on the uncertainty of the instances, the model learns an action strategy that determines when to predict or request the label of each instance. Through the experiments on Omniglot dataset, the model shows its ability to achieve a good prediction accuracy with few label requests.**

Keywords-active learning; meta learning; reinforcement learning; GRU.

I. INTRODUCTION

Active learning [1] uses unlabeled and labeled instances to train a highly accurate classifier to reduce the workload of human experts. The algorithm simulates the human learning process, selects part of instances to label and iteratively improves the generalization performance of the classifier. Therefore, it has been widely used in information retrieval and text, image and speech recognition in recent years.

Most of the traditional active learning methods are carefully formulating some criteria for selecting samples, such as uncertainty sampling [2], query-by-committee [3], margin [4] and representative and diversity-based sampling [5]. It's hard to pinpoint which approach is better, because each approach starts from a reasonable, meaningful and completely different motivation. However, for now, there is no universal criteria that performs well on all datasets. This paper introduces a learning-based approach, rather than a manually-designed sample-selection criterion, which integrates active learning algorithm with reinforcement learning. Our approach not only learns to use small supervisors to classify instances, but also learns about label-querying strategies. The model adopts a stream-based active learner that considers the online environment for active learning.

Our primary contribution in this work is using a GRU to improve the active one-shot learning model introduced by Woodward *et al.* [6]. We evaluate the model on Omniglot ("active" variants of existing one-shot learning tasks [7]), and our experiment results show that it can learn label-querying strategies efficiently with simpler structure.

The rest of this paper is structured as follows. Section II summarizes the existing approaches related to our work. Section III presents the task and the general framework of our proposed active learning model. Section IV presents the experiments and interprets the results. Finally, we conclude this paper in Section V.

II. RELATED WORK

Active learning has been well studied in the past few decades. The main idea of the active learning is that a learner should achieve higher accuracy with fewer labeled training instances, if it is able to choose the training instances from which it learns. Numerous algorithms have been proposed to design the criteria for the selection of which examples to label [2][5][8][9]. However, most of these traditional active learning methods are based on heuristics, which may be limited when the data distribution of the underlying learning problems vary (e.g. a new class appears). Instead, we used a meta-learning approach to train an active learner via reinforcement learning to solve a one-shot learning task. The idea of combing active learning and reinforcement learning was recently investigated by Woodward *et al.* [6]. In contrast to their work, we used a GRU instead of a Long Short-Term Memory (LSTM) network to approximate the action-value function in reinforcement learning. Compared with LSTM, GRU has fewer parameters, so it can effectively speed up the training process [10] and requires fewer samples, which is more suitable for our one-shot active learning task. Similar inspirations have also been studied by Bachman *et al.* [7] Pang *et al.* [11] and Puzanov *et al.* [12].

 ISBN: 978-1-61208-698-9

III. MODEL DESCRIPTION

A. Task Description

We mainly focus on the stream-based online active learning scenario [6], in which instances can be continually obtained from the data stream and presented in an exogenously-determined order. Thus, the input of the model is a stream of images, the label of which is queried or predicted by our model. The performance of our model was improved with a short training episode and a small number of examples per class to maximize the performance of test episodes, which consists of classes that are not encountered in training. The structure of our task can be seen in Figure 1. The classes and their labels and the specific samples are shuffled and randomly presented at each episode.

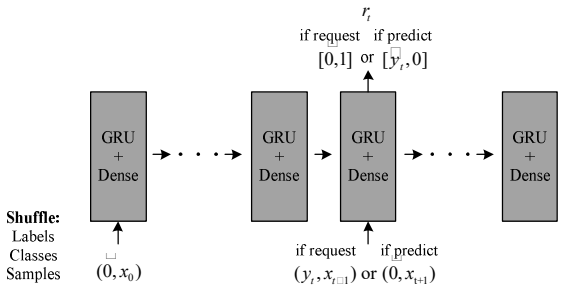


Figure 1. Task structure

Figure 2. Example of 3-way problem instance on Omniglot.

At each time step, the input of the model is an image along with a vector that depends on the output taken previous instance as input. The *N*-way task is set up as follows: pick *N* unseen classes per episode. Figure 2 shows an example of a 3-way problem on Omniglot. The output of the model is a one-hot vector of length *N +1*. If the model requests the label of the image x_t, it sets the final bit of the output vector of this timestep to 1, which means the output of timestep t is $[\vec{0}, 1]$. Thus, the reward of this label request action is R_{req}. The true label y_t of image x_t is then provided at the next time step along with the next image x_{t+1}, so the input of $t+1$ is (y_t, x_{t+1}). Alternatively, if the model makes a prediction of x_t, it sets one of the first k bits of the output vector to represent $\hat{y}$, so the output of this step is $[\hat{y}, 0]$. The reward of this action is R_{cor} if the prediction is correct or R_{inc} if incorrect. If a prediction is made at time step t, no information of its true label y_t is supplied at the next time step $t+1$, then the input is $(\vec{0}, x_{t+1})$ instead.

B. Methodology

We use a model-free reinforcement learning method Q-learning to learn an optimal action strategy, which can maximize the rewards. The loss function we use is defined as,

$$\mathcal{L}(\theta) \coloneqq \sum_t \Big[Q(o_t, a_t) - \Big(r_t + \gamma \max_{a_{t+1} \in \mathcal{A}} Q^*(s_{t+1}, a_{t+1}) \Big) \Big]^2 \quad (1)$$

where, θ are the parameters of the function approximator, o_t are the observations such as images that the agent receives, a_t is the action the model chooses at timestep t, Q^* is the optimal value of action-value function Q.

We use a GRU [13] network connected to a linear output layer to adopt the methodology of using action-value function $Q(o_t, a_t)$ in Q-learning. $Q(o_t)$ outputs a vector, where each element corresponds to an action:

$$Q(o_t, a_t) = Q(o_t) \cdot a_t \quad (2)$$

$$Q(o_t) = W^{hq} h_t + b^q \quad (3)$$

where, b^q is the action-value bias, h_t is the output of the GRU, W^{hq} are the weights mapping from the GRU output to action-values.

IV. EXPERIMENTS

A. Setup

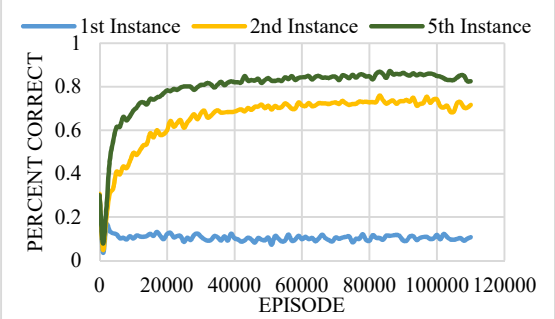
The experiments were carried out on the Omniglot dataset [14] that contains 32460 instances having 1623 classes of characters from 50 different alphabets, each hand-written by 20 different persons. The dataset was randomly divided into 1200 characters for training and the rest 423 characters are kept for testing. The images were downscaled to 28×28 pixels and each pixel was normalized between 0.0 and 1.0.

30 Omniglot images from 3 random classes were chosen in each episode. Each class of images was randomly rotated in $\{0°, 90°, 180°, 270°\}$. A GRU with 200 hidden units was used here. We optimized the parameters of our model using Adam with the default parameters [15]. A grid search was performed over the following parameters, and the parameters of the results reported in this article are listed here. During the training process, epsilon-greedy ($\epsilon = 0.23$) exploration is set for actions selection. The learning rate of training was set to 0.001 and the discount factor γ was set to 0.8. The reward values were set as: $R_{cor} = +1, R_{inc} = -1$, $R_{req} = -0.3$.

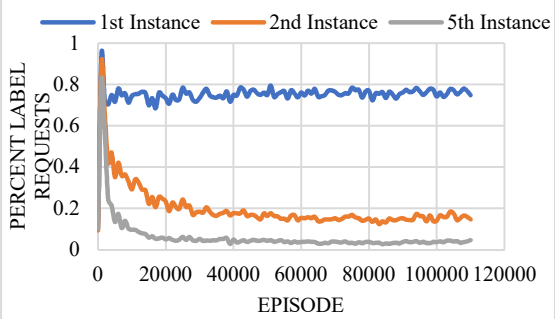
B. Results

Here, we present the results of our experiments. The 1st, 2nd, 5th of all classes in each episode were identified. After 100,000 episodes, training is ceased and the model was given 10,000 more test episodes. No further updates occurred during these episodes. The results can be seen in Figure 3.

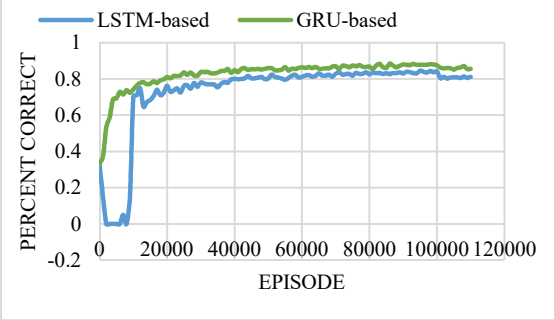
 ISBN: 978-1-61208-698-9



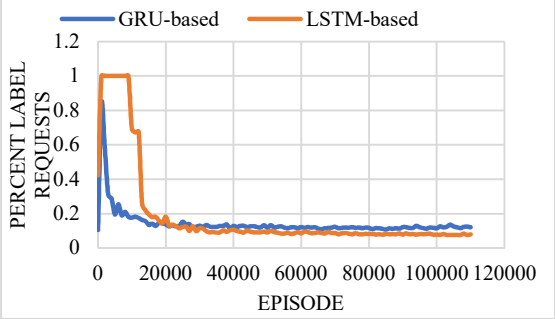
(a) Identified accuracy



(b) Identified label request rate



(c) Overall accuracy



(d) Overall request rate

Figure 3. Experiment results

As can be seen in the plot, the proposed GRU-based meta-learning model learns to query the label for early instances of a class, and makes more prediction for later instances. Simultaneously, the accuracy of the model is improved on later instances of a class. It shows that our model has learned an effective querying strategy that effectively requests tags when new classes are present, and quickly learns useful information to make accurate predictions when they encounter the same category in the future. After initial training, our model accuracy rate was stable at 85%, while the label request rate was stable at 12%. Compared with supervised learning, our model greatly reduces the dependence on the number of labels and human workload, and achieves decent prediction accuracy. At the same time, our method speeds up the convergence of the algorithm compared to the LSTM-based method [6].

V. CONCLUSION

We introduced a GRU-based meta-learning model that learns active learning in an reinforcement learning way and experimented it on Omniglot one-shot learning tasks. Our results show that our model can learn an optimal query strategy and achieve a good classification accuracy with a small amount of labeled data.

As we used a GRU network to approximate the action-value function in reinforcement learning, a promising direction is that the GRU network can be replaced by a more sophisticated one-shot learning approach such as Matching Network [16] or Memory-Augmented Neural Networks [17]. We will leave this as our future work.

REFERENCES

[1] B. Settles, “Active Learning Literature Survey,” University of Wisconsinmadison, vol. 39, no. 2, pp. 127–131, 2009.

[2] A. Kapoor, K. Grauman, R. Urtasun, and T. Darrell, “Active Learning with Gaussian Processes for Object Categorization,” vol. 88, no. 2, pp. 1-8, 2015.

[3] Seung, S. H., Opper, and Sompolinsky, “Query by committee,” Proc of the Fith Workshop on Computational Learning Theory, vol. 284, pp. 287-294, 1992.

[4] S. Tong, and D. Koller, Support vector machine active learning with applications to text classification: JMLR.org, 2002.

[5] R. Chattopadhyay, Z. Wang, W. Fan, I. Davidson, S. Panchanathan, and J. Ye, “Batch Mode Active Sampling based on Marginal Probability Distribution Matching,” Acm Transactions on Knowledge Discovery from Data, vol. 7, no. 3, pp. 1-25, 2013.

[6] M. Woodward, and C. Finn, “Active One-shot Learning,” 2017.

[7] P. Bachman, A. Sordoni, and A. Trischler, “Learning Algorithms for Active Learning,” 2017.

[8] S. Huang, J. Chen, X. Mu, and Z. Zhou, "Cost-Effective Active Learning from Diverse Labelers." pp. 1879-1885.

[9] R. Giladbachrach, A. Navot, and N. Tishby, "Query by Committee Made Real." pp. 443-450.

[10] J. Chung, C. Gulcehre, K. H. Cho, and Y. Bengio, “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling,” Eprint Arxiv, 2014.

[11] K. Pang, M. Dong, Y. Wu, and T. Hospedales, “Meta-Learning Transferable Active Learning Policies by Deep Reinforcement Learning,” 2018.

[12] A. Puzanov, and K. Cohen, “Deep Reinforcement One-Shot Learning for Artificially Intelligent Classification Systems,” 2018.

[13] L. Agatha, D. Arnaud, P. D. Walshaw, A. L. Cho, R. M. Bilder, J. J. Mcgough, J. T. Mccracken, M. Scott, and S. K. Loo, “Electroencephalography correlates of spatial working memory deficits in attention-deficit/hyperactivity disorder: vigilance, encoding, and maintenance,” Journal of Neuroscience the Official Journal of the Society for Neuroscience, vol. 34, no. 4, pp. 1171-82, 2014.

[14] B. M. Lake, R. Salakhutdinov, and J. B. Tenenbaum, “Human-level concept learning through probabilistic program induction,” Science, vol. 350, no. 6266, pp. 1332-1338, 2015.

[15] D. P. Kingma, and J. Ba, “Adam: A Method for Stochastic Optimization,” Computer Science, 2014.

[16] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra, “Matching Networks for One Shot Learning,” 2016.

[17] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. Lillicrap, “One-shot Learning with Memory-Augmented Neural Networks,” 2016.

 ISBN: 978-1-61208-698-9