One-Shot Learning using Mixture of Variational Autoencoders: a Generalization Learning approach

Extended Abstract*

Decebal Constantin Mocanu
Eindhoven University of Technology
Eindhoven, Netherlands
d.c.mocanu@tue.nl

Elena Mocanu
Eindhoven University of Technology
Eindhoven, Netherlands
e.mocanu@tue.nl

ABSTRACT

Deep learning, even if it is very successful nowadays, traditionally needs very large amounts of labeled data to perform excellent on the classification task. In an attempt to solve this problem, the one-shot learning paradigm, which makes use of just one labeled sample per class and prior knowledge, becomes increasingly important. In this paper, we propose a new one-shot learning method, dubbed MoVAE (Mixture of Variational AutoEncoders), to perform classification. Complementary to prior studies, MoVAE represents a shift of paradigm in comparison with the usual one-shot learning methods, as it does not use any prior knowledge. Instead, it starts from zero knowledge and one labeled sample per class. Afterward, by using unlabeled data and the generalization learning concept (in a way, more as humans do), it is capable to gradually improve by itself its performance. Even more, if there are no unlabeled data available MoVAE can still perform well in one-shot learning classification. We demonstrate empirically the efficiency of our proposed approach on three datasets, i.e. the handwritten digits (MNIST), fashion products (Fashion-MNIST), and handwritten characters (Omniglot), showing that MoVAE outperforms state-of-the-art one-shot learning algorithms.

KEYWORDS

One-Shot Learning; Semi-Supervised Learning; Variational Autoencoders; Generalization Learning; Collective Intelligence

ACM Reference Format:
Decebal Constantin Mocanu and Elena Mocanu. 2018. One-Shot Learning using Mixture of Variational Autoencoders: a Generalization Learning approach. In *Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), Stockholm, Sweden, July 10–15, 2018,* IFAAMAS, 3 pages.

1 INTRODUCTION

Object recognition is an important problem, and it has many applications, e.g. computer vision [1, 3, 12, 13], robotics [11] and healthcare [10]. Traditional solutions use classifiers built on large amounts of data. In a time with more and more unlabeled data, manually labeling of all these data is costly, time consuming, and inefficient. Hence, the one-shot learning paradigm becomes increasingly important. The aims of this paradigm is to improve the generalization capabilities of the learning models in such a way that they are capable to achieve a very good performance by using just one labeled sample per class or (at maximum) few labeled samples. To achieve this, usually the state-of-the-art one-shot learning algorithms make use of prior knowledge and large amounts of unlabeled data. Even so, up to our best knowledge, the maximum classification accuracy achieved, for instance, on MNIST by one-shot learning algorithms with one labeled sample per class (1-shot) is just about 72%.

In this paper, we address the above problem, and we propose a new one-shot learning classification method, dubbed Mixture of Variational Autoencoders (MoVAE). Contrary to the state-of-the-art one-shot learning methods, MoVAE does not need at all any prior knowledge. In fact, it complements these methods. It starts from zero knowledge and one (or few) labeled samples per class, and then it gradually learns to generalize its knowledge using the generalization learning concept [15]. Also, by opposite to the usual direction in artificial neural networks, MoVAE is not an unitary neural network. In fact, it is composed by many Variational Autoencoders (VAEs) [6], each one learning the distribution of a class. Thus, MoVAE can be a good example of collective intelligence. Each VAE took separately can not perform classification, but all of them acting together, are able to learn and classify objects very well.

2 MIXTURE OF VARIATIONAL AUTOENCODERS

The intuition behind MoVAE is simple and it is inspired by human learning processes. People, when they learn new concepts, they do not manage too well to deal with large amounts of labeled data, but they are often extremely efficient to generalize across various conditions just from one example. Sometimes, they make use of prior acquired knowledge, and sometimes not. They start just from one example and gradually add new representations of that example (or situation) to its default category using generalization [4]. At a different scale, the learning concept evolved through the human world into a collective intelligence behavior. The advances of human society were mainly made, not by super-humans, but by many humans, connected between them in a social network, sharing a set of values, and working together for a common goal. Moreover, a human is far to be one of the strongest animal in the world. In fact, it is quite weak, but humans collaborative way of being and personal specialization made from the human race one of the most successful in the word [5].

*The full version of this paper is available at:
https://arxiv.org/abs/1804.07645
https://pure.tue.nl/ws/portalfiles/portal/95923754

Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), M. Dastani, G. Sukthankar, E. André, S. Koenig (eds.), July 10–15, 2018, Stockholm, Sweden.

Keeping the proportion, by analogy, we argue that in machine learning, we should not search for the most powerful model possible, but to create many specialized models, each being capable of doing well its specialized task. Then, these models working together will be able to fulfill a common goal, inaccessible for a singular model. In a way, in artificial intelligence, this approach is followed by ensembles and swarm intelligence, with the difference that each particle or ensemble could do a better or worse job on the common task, while in what we propose next, one singular model would achieve nothing.

These being said, and knowing that a Variational Autoencoder can represent very well a data distribution, in this paper, we propose to build a Mixture of Variational Autoenconders (MoVAE) to perform classification. In the specific case of classification, each VAE of the mixture will be very specialized and will learn the distribution of just one class by being trained on samples belonging just to its specific class. Thus, after the learning phase, our assumption is that each VAE will reconstruct very well unseen images belonging to its encoded class, but if images belonging to other classes will be reconstructed through it, then their reconstructed version will be not so good. And here comes the trick of cooperative inference. Each VAE model is not able to discriminate if a given image belongs to its encoded class if it looks just of that image reconstructed version, but the mixture of VAEs it is. If we pass the same image through all the VAEs belonging to the mixture then we obtained a reconstructed version of the original image for any VAE. Then the class of the original image is given by the VAE which obtains the best reconstruction of the original image. Moreover, our assumption is that our proposed approach does not need many labeled images to learn well the class distributions. In fact, it can use just one labeled sample per class to encode in a decent manner the corresponding class in each VAE. Then, by using generalization learning and considering unlabeled data it will be able to gradually increase the quality of the encoded distributions, being capable to improve by itself its discriminative capabilities, as described in the full version of this paper.

3 EXPERIMENTS AND RESULTS

Herein, we briefly report MoVAE performance on a small set of experiments, while the interested reader is referred to the full version of this paper for a thorough analyze and more scenarios on the MNIST [9], Fashion-MNIST [17], and Omniglot [8] datasets.

One-shot semi-supervised learning. In this set of experiments, we have evaluated MoVAE on the MNIST and Fashion-MNIST datasets. We consider just 1, 5, and 10 randomly chosen labeled samples per class (from here comes the *one-shot learning* part of the paragraph name). All the other samples belonging to the training sets were used as unlabeled data (from here comes the *semi-supervised learning* part of the paragraph name). Table 1 presents the results. We may observe that on both datasets, MoVAE models achieve good accuracies, outperforming the ones obtained by the Convolutional Neural Network (CNN) models offered as an example in the Keras library [2]. While Fashion-MNIST dataset is very new and not too many one-shot learning results are reported yet in the literature, it can be observed that on the MNIST dataset MoVAE outperforms clearly the state-of-the-art machine learning models.

Table 1: One-shot semi-supervised learning - Classification accuracy of MoVAE against baseline CNN and state-of-the-art using 1, 5, and 10 labeled samples per class on the MNIST and Fashion-MNIST datasets.

Model	Labeled samples/ class [#]	Data Augmen-tation	Prior Knowledge	Unlabeled Data	MNIST Accuracy [%]	Fashion-MNIST Accuracy [%]
MoVAE (ours)	1-shot	no	no	yes	69.6±6.5	-
MoVAE (ours)	1-shot	yes	no	yes	**91.1±4.7**	**61.6±2.8**
CNN	1-shot	no	no	no	17.4±3.5	-
CNN	1-shot	yes	no	no	22.1±3.4	21.3±4.3
CPM [16]	1-shot	-	yes	no	68.8	-
Siamese Net [7]	1-shot	-	yes	no	70.3	-
Matching Nets [14]	1-shot	-	yes	no	72.0	-
MoVAE (ours)	5-shot	no	no	yes	90.4±1.6	-
MoVAE (ours)	5-shot	yes	no	yes	**94.5±0.6**	**66.5±1.7**
CNN	5-shot	no	no	no	24.3±5.4	-
CNN	5-shot	yes	no	no	28.1±5.2	28.2±4.7
CPM [16]	5-shot	-	yes	no	83.8	-
MoVAE (ours)	10-shot	no	no	yes	93.1±1.1	-
MoVAE (ours)	10-shot	yes	no	yes	**94.9±0.4**	**70.5±1.9**
CNN	10-shot	no	no	no	33.1±5.1	-
CNN	10-shot	yes	no	no	47.7±6.6	36.6±5.4
CPM [16]	10-shot	-	yes	no	≈88.0	-

Table 2: One-shot learning - MoVAE performance on the 1623-way Omniglot (a new challenge for one-shot learning).

Model	Labeled samples/ class [#]	Data Augmen-tation	Prior Knowledge	Unlabeled Data	1623-way Omniglot Accuracy [%]
MoVAE (ours)	1-shot	yes	no	no	**27.8±0.4**
kNN	1-shot	yes	no	no	3.1±0.01
Random guess	-	-	-	-	0.06
MoVAE (ours)	5-shot	yes	no	no	**43.2±0.1**
kNN	5-shot	yes	no	no	5.9±0.01
Random guess	-	-	-	-	0.06

One-shot learning. Herein, we have addressed a pure one-shot learning problem by performing 1623-way (1623-classes) one-shot classification on the Omniglot dataset [8]. We used, randomly chosen, 1 (1-shot) and 5 (5-shot) labeled samples per class (character). The remaining samples belonging to the same class (19 and 15 samples, respectively) were used as testing data. We did not consider at all unlabeled data, and we did not used the generalization learning capabilities of MoVAE. Table 2 reports the results. Please note that Siamese Net [7] and Matching Nets [14] methods are not capable of performing 1623-way classification on Omniglot as they need all the data from some classes to create the prior knowledge. An exception for this situation would be the use of other datasets to create the prior knowledge. However, to the best of our knowledge, there are no results reported in the literature for this situation.

4 CONCLUSION

In this paper, we introduce MoVAE (Mixture of Variational Autoencoders), taking inspiration from the human world. MoVAE is capable to successfully perform the one-shot learning task, without the need of having prior knowledge, due to its generalization learning capabilities. Even when unlabeled data is unavailable, MoVAE offers good performance. Thus, we introduce 1623-way 1-shot learning classification on Omniglot, a *new challenge for one-shot learning*. Herein, MoVAE accuracy is 463 times higher than the one of random guess and 9 times higher than the one of kNN, while no other state-of-the-art results are reported in this very difficult context.

REFERENCES

[1] P. Agrawal, J. Carreira, and J. Malik. 2015. Learning to See by Moving. In *2015 IEEE International Conference on Computer Vision (ICCV)*. 37–45. https://doi.org/10.1109/ICCV.2015.13

[2] François Chollet. 2015. keras. https://github.com/fchollet/keras. (2015).

[3] N. Dalal and B. Triggs. 2005. Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 1. 886–893 vol. 1. https://doi.org/10.1109/CVPR.2005.177

[4] Mark A. Gluck, Eduardo Mercado, and Catherine E. Myers. 2011. *Learning and Memory: From Brain to Behavior* (2nd ed.). New York: Worth Publishers.

[5] Yuval Noah Harari. 2015. *Sapiens: A Brief History of Humankind.*

[6] D. P. Kingma and M. Welling. 2013. Auto-encoding variational Bayes. *CoRR* arXiv:1312.6114 (2013).

[7] Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. 2015. Siamese Neural Networks for One-shot Image Recognition.

[8] Brenden M. Lake, Ruslan Salakhutdinov, and Joshua B. Tenenbaum. 2015. Human-level concept learning through probabilistic program induction. *Science* 350, 6266 (11 Dec. 2015), 1332–1338. https://doi.org/10.1126/science.aab3050

[9] Yann LeCun, LÃĺon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-Based Learning Applied to Document Recognition. In *Proceedings of the IEEE*, Vol. 86. 2278–2324.

[10] Geert J. S. Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A. W. M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. 2017. A Survey on Deep Learning in Medical Image Analysis. *CoRR* abs/1702.05747 (2017). http://arxiv.org/abs/1702.05747

[11] Patricio Loncomilla, Javier Ruiz del Solar, and Luz MartÃŋnez. 2016. Object recognition using local invariant features for robotic applications: A survey. *Pattern Recognition* 60 (2016), 499 – 514. https://doi.org/10.1016/j.patcog.2016.05.021

[12] W. Ouyang, X. Wang, X. Zeng, Shi Qiu, P. Luo, Y. Tian, H. Li, Shuo Yang, Zhe Wang, Chen-Change Loy, and X. Tang. 2015. DeepID-Net: Deformable deep convolutional neural networks for object detection. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2403–2412. https://doi.org/10.1109/CVPR.2015.7298854

[13] J. Thewlis, H. Bilen, and A. Vedaldi. 2017. Unsupervised object learning from dense invariant image labelling. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*.

[14] Oriol Vinyals, Charles Blundell, Timothy P. Lillicrap, Koray Kavukcuoglu, and Daan Wierstra. 2016. Matching Networks for One Shot Learning. (2016), 3630–3638.

[15] D.A. Waterman. 1970. Generalization learning techniques for automating the learning of heuristics. *Artificial Intelligence* 1, 1 (1970), 121 – 170.

[16] Alex Wong and Alan L Yuille. 2015. One shot learning via compositions of meaningful patches. In *Proceedings of the IEEE International Conference on Computer Vision*. 1197–1205.

[17] Han Xiao, Kashif Rasul, and Roland Vollgraf. 2017. Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms. (2017).