Challenges in Representation Learning: A Report on Three Machine Learning Contests

Ian J. Goodfellow[1], Dumitru Erhan[2], Pierre Luc Carrier, Aaron Courville, Mehdi Mirza, Ben Hamner, Will Cukierski, Yichuan Tang, David Thaler, Dong-Hyun Lee, Yingbo Zhou, Chetan Ramaiah, Fangxiang Feng, Ruifan Li, Xiaojie Wang, Dimitris Athanasakis, John Shawe-Taylor, Maxim Milakov, John Park, Radu Ionescu, Marius Popescu, Cristian Grozea, James Bergstra, Jingjing Xie, Lukasz Romaszko, Bing Xu, Zhang Chuang, and Yoshua Bengio

[1] Université de Montréal, Montréal QC H3T 1N8, Canada
goodfeli@iro.umontreal.ca
[2] Google, Venice, CA 90291, USA
dumitru@google.com

Abstract. The ICML 2013 Workshop on Challenges in Representation Learning[1] focused on three challenges: the black box learning challenge, the facial expression recognition challenge, and the multimodal learning challenge. We describe the datasets created for these challenges and summarize the results of the competitions. We provide suggestions for organizers of future challenges and some comments on what kind of knowledge can be gained from machine learning competitions.

Keywords: representation learning, competition, dataset.

1 Introduction

This paper describes three machine learning contests that were held as part of the ICML workshop "Challenges in Representation Learning." The purpose of the workshop, organized by Ian Goodfellow, Dumitru Erhan, and Yoshua Bengio, was to explore the latest developments in representation learning, with a special emphasis on testing the capabilities of current representation learning algorithms (See [1] for a recent review) and pushing the field towards new developments via these contests. Ben Hamner and Will Cukierski handled all issues related to Kaggle hosting and ensured that the contests ran smoothly. Ian Goodfellow and Dumitru Erhan provided baseline solutions to each challenge, mostly in Pylearn2 [2] format. Google provided prizes for all three contests. The winner of each contest received $350 while the runner-up received $150. A diverse range of competitors spanning academia, industry, and amateur machine learning provided excellent solutions to all three problems. In this paper, we summarize their solutions, and discuss what we can learn from them.

[1] http://deeplearning.net/icml2013-workshop-competition

M. Lee et al. (Eds.): ICONIP 2013, Part III, LNCS 8228, pp. 117–124, 2013.

2 The Black Box Learning Challenge

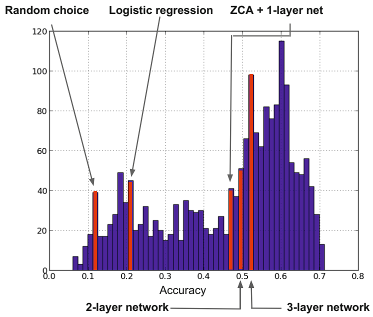

Fig. 1. Histogram of accuracies obtained by different submissions on the BBL-2013 dataset. Organizer-provided baselines shown in red.

The black box learning challenge[2] was designed with two goals in mind. First, the data was obfuscated, so that competitors could not use human-in-the-loop techniques like visualizing filters to guide algorithmic development. A common criticism of deep learning is that it is an art requiring an expert practitioner. By keeping the domain of the data secret, this contest reduced the usefulness of the human practitioner. This idea was similar to a recent DARPA-organized unsupervised and transfer learning challenge [3] which used obfuscated data and required submission of a representation of the data that would then be used on the competition server to train a very weak classifier. In this contest, we allowed competitors to use any method; using representation learning was not a requirement. The second goal of this contest was to test the ability of algorithms to benefit from extra unsupervised data. To this end, we provided only very few labeled examples.

This contest introduced the Black Box Learning 2013 (BBL-2013) dataset. The scripts needed to re-generate it are available for download[3]. The dataset is an obfuscated subset of the second (MNIST-like) format of the Street View House Numbers dataset[4]. Dumitru Erhan created the dataset. The original data contained 3,072 features (pixels) which he projected down to 1875 by multiplication by a random matrix. He also removed one class (the "4"s). These measures obfuscated the data so competitors did not know what task they were solving. The organizers did not reveal the source of the dataset until after the contest was over. To make the challenge emphasize semi-supervised learning, only 1,000 labeled examples were kept for training. Another 5,000 were used for the public leaderboard. For these examples, the labels are not provided to the competitors, but the features are. Each team may upload predictions for these examples twice per day. The resulting accuracy is published publicy. The public test set is thus a sort of validation set, but also gives one's competitors information. Another 5,000 examples were used for the private test set. The features for these examples are given to the competitors as well, but only the contest administrators see the accuracy on them until after the contest has ended. The

[2] http://www.kaggle.com/c/challenges-in-representation-learning-the-black-box-learning-challenge
[3] http://www-etud.iro.umontreal.ca/~goodfeli/bbl2013.html