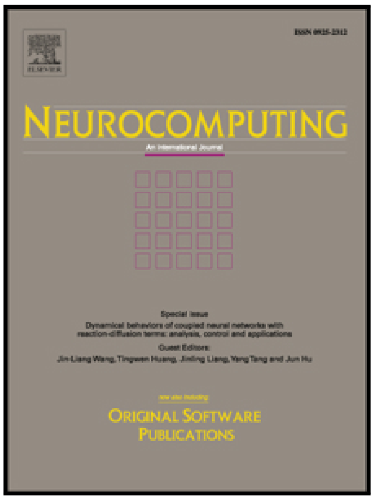
Accepted Manuscript



Bayesian Optimization on Graph-Structured Search Spaces: Optimizing Deep Multimodal Fusion Architectures

Dhanesh Ramachandram, Michal Lisicki, Timothy J. Shields, Mohamed R. Amer, Graham W. Taylor

PII:	S0925-2312(18)30212-1
DOI:	10.1016/j.neucom.2017.11.071
Reference:	NEUCOM 19356

To appear in:	*Neurocomputing*

Received date:	10 July 2017
Revised date:	31 October 2017
Accepted date:	20 November 2017

Please cite this article as: Dhanesh Ramachandram, Michal Lisicki, Timothy J. Shields, Mohamed R. Amer, Graham W. Taylor, Bayesian Optimization on Graph-Structured Search Spaces: Optimizing Deep Multimodal Fusion Architectures, *Neurocomputing* (2018), doi: 10.1016/j.neucom.2017.11.071

Bayesian Optimization on Graph-Structured Search Spaces: Optimizing Deep Multimodal Fusion Architectures

Dhanesh Ramachandram[†a], Michal Lisicki[†a], Timothy J. Shields[†b], Mohamed R. Amer[b], Graham W. Taylor[a,c,*]

[a]*Machine Learning Research Group, School of Engineering, University of Guelph, Canada.*
[b]*Center for Vision Technologies, SRI International, USA.*
[c]*Canadian Institute for Advanced Research.*

Abstract

A popular testbed for deep learning has been multimodal recognition of human activity or gesture involving diverse inputs like video, audio, skeletal pose and depth images. Deep learning architectures have excelled on such problems due to their ability to combine modality representations at different levels of nonlinear feature extraction. However, designing an optimal architecture in which to fuse such learned representations has largely been a non-trivial human engineering effort. We treat fusion structure optimization as a hyperparameter search and cast it as a discrete optimization problem under the Bayesian optimization framework. We propose two methods to compute structural similarities in the search space of tree-structured multimodal architectures, and demonstrate their effectiveness on two challenging multimodal human activity recognition problems.

Keywords: Deep learning, Bayesian optimization, kernel methods

1. Introduction

With the increasing complexity of deep architectures (e.g. Szegedy et al., 2015; He et al., 2016) finding the right architecture and associated hyperparameters, in a process known as *model search*, has kept humans "in-the-loop." A practitioner may train and evaluate models millions of times to find a set of hyperparameters that lead to optimal performance on a given dataset. Traditionally, the deep learning community has employed systematic search methods, such as grid search and random search (Bergstra and Bengio, 2012). In recent years, model-based search, and in particular, Bayesian Optimization (BO), has become the preferred technique in many deep learning applications (Shahriari et al., 2016).

It is common to apply BO to the search over various architectural hyperparameters (e.g. number of layers, number of hidden units per layer), but applying BO to search the space of complete network architectures is much more challenging. Modeling the space of network architectures as a discrete topological space is equivalent to making random choices between architectures during the BO procedure, where each architecture variant in the search space would be equally similar. Therefore, we must define some distance metric between architectures in order to exploit architectural similarities.

[†]Authors contributed equally.
[*]Corresponding author
Email address: gwtaylor@uoguelph.ca (Graham W. Taylor)