Available online at www.sciencedirect.com

ScienceDirect

Mathematics and Computers in Simulation 82 (2011) 442–449

MATHEMATICS AND COMPUTERS IN SIMULATION

www.elsevier.com/locate/matcom

Original article

Learning over sets with Recurrent Neural Networks: An empirical categorization of aggregation functions

W. Heidl[a,*], C. Eitzinger[a], M. Gyimesi[b], F. Breitenecker[b]

[a] Profactor GmbH, Im Stadtgut A2, 4407 Steyr-Gleink, Austria
[b] Vienna University of Technology, Wiedner Hauptstraße 8-10, 1040 Wien, Austria

Received 4 February 2010; accepted 14 October 2010
Available online 26 October 2010

Abstract

Numerous applications benefit from parts-based representations resulting in sets of feature vectors. To apply standard machine learning methods, these sets of varying cardinality need to be aggregated into a single fixed-length vector. We have evaluated three common Recurrent Neural Network (RNN) architectures, Elman, Williams & Zipser and Long Short Term Memory networks, on approximating eight aggregation functions of varying complexity. The goal is to establish baseline results showing whether existing RNNs can be applied to learn order invariant aggregation functions. The results indicate that the aggregation functions can be categorized according to whether they entail (a) *selection* of a subset of elements and/or (b) *non-linear* operations on the elements. We have found that RNNs can very well learn to approximate aggregation functions requiring either (a) or (b) and those requiring only linear sub functions with very high accuracy. However, the combination of (a) and (b) cannot be learned adequately by these RNN architectures, regardless of size and architecture.

Keywords: Recurrent Neural Networks; Order invariance; Aggregation

1. Introduction

Numerous applications benefit from parts-based representations resulting in sets of feature vectors. As an example, the good/bad decision in surface inspection tasks often depends on the whole set of extracted fault feature vectors, such as their spatial distribution, their total area and similar quantities. This information cannot be utilized if faults are processed one-by-one. Learning the decision process in this context thus requires methods for classification of sets of feature vectors extracted from the faults [11].

To apply standard machine learning methods, these sets of varying cardinality need to be aggregated into a single fixed-length vector. We define the classification task over sets $X_i{:=}\{x_{i,1}, x_{i,2}, \ldots, x_{i,n_i}\}$ of feature vectors $x_{i,j} \in \mathbb{R}^m$

$$f \circ g : \mathbb{R}^{m \cdot n_i} \mapsto \{-1, 1\} \tag{1}$$

$$X_i \mapsto c_i{:=}(f \circ g(X_i)) \tag{2}$$

[*] Corresponding author. Tel.: +43 7252 885 252; fax: +43 7252 885 101.
E-mail addresses: wolfgang.heidl@profactor.at (W. Heidl), christian.eitzinger@profactor.at (C. Eitzinger), mgyimesi@osiris.tuwien.ac.at (M. Gyimesi), fbreiten@osiris.tuwien.ac.at (F. Breitenecker).

0378-4754/$36.00 © 2010 IMACS. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.matcom.2010.10.018