Inclusion of Symbolic Domain-Knowledge into Deep Neural Networks

THESIS

Submitted in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

by

Tirtharaj Dash
(ID No. 2016PHXF0421G)

Under the Supervision of
Ashwin Srinivasan

and

Co-supervision of
Sukanta Mondal

COMPUTER SCIENCE AND INFORMATION SYSTEMS
BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE
PILANI – 333 031 (INDIA)
April 2022

Abstract

This dissertation is concerned with techniques for inclusion of domain-knowledge into Deep Neural Networks (DNNs). We are primarily concerned with real-world scientific problems with the following characteristics: (a) Data are naturally graph-structured (relational), (b) The amount of data available is typically small, and (c) There is significant domain-knowledge, usually expressed in some logical form (rules, taxonomies, constraints and the like). Broadly, there are 3 different ways in which the domain-knowledge can be incorporated into a DNN: by changing the input representation, by changing the loss function, or by changing the model (structure and parameters). We propose techniques for the inclusion of domain-knowledge into DNNs that change the input representation. In particular, our principal contributions are as follows: (1) We study the inclusion of complex domain-knowledge into Multilayer Perceptrons (MLPs) using relational features and propositionalisation. We propose a utility-based stochastic sampling technique for drawing features from a large but countable space of relational features; (2) We propose a simplified technique called 'vertex-enrichment' for incorporating symbolic domain knowledge into deep neural networks that deal with graph-structured data, known as graph neural networks (GNNs); (3) We propose a systematic technique to incorporate symbolic domain-knowledge into GNNs using the method of inverse entailment available in Inductive Logic Programming (ILP); and (4) We construct a sequence generation system using a modular combination of two deep generative models and a discriminator model based on (3), and use this system for a problem of early-stage lead discovery in drug design. Our implementations are techniques that combine neural networks and symbolic representations, resulting in new neuro-symbolic models, such as: Deep Relational Machines (DRMs), Vertex-Enriched Graph Neural Networks (VEGNNs), Bottom-Graph Neural Networks (BotGNNs), and a modular end-to-end neuro-symbolic system for the generation of novel molecules for drug design. Our primary hypothesis is that inclusion of domain-knowledge can significantly improve the performance of a deep neural network. We conduct large-scale empirical testing of our hypothesis, using nearly 75 datasets in the broad area of drug discovery that consist of over $200,000$ relational data instances and with domain-knowledge containing about 100 relations. In all cases, our empirical evidence supports the primary hypothesis and encourages the inclusion of domain-knowledge into deep neural networks for prediction and explanation.

List of Publications

The work carried out in this dissertation have appeared in the following peer-reviewed publications in reverse chronological order:

1. T. Dash, S. Chitlangia, A. Ahuja, A. Srinivasan, "A review of some techniques for inclusion of domain-knowledge into deep neural networks", *Nature Scientific Reports*, 2022.
   URL: https://doi.org/10.1038/s41598-021-04590-0

2. T. Dash, A. Srinivasan, L. Vig, A. Roy, "Using domain-knowledge to assist lead discovery in early-stage drug design", *International Conference on Inductive Logic Programming*, 2021.
   URL: https://doi.org/10.1007/978-3-030-97454-1_6

3. T. Dash, A. Srinivasan, A. Baskar, "Inclusion of domain-knowledge into GNNs using mode-directed inverse entailment", *Machine Learning*, 2021.
   URL: https://doi.org/10.1007/s10994-021-06090-8

4. T. Dash, A. Srinivasan, L. Vig, "Incorporating symbolic domain knowledge into graph neural networks", *Machine Learning*, 2021.
   URL: https://doi.org/10.1007/s10994-021-05966-z

5. T. Dash, A. Srinivasan, R.S. Joshi, A. Baskar, "Discrete stochastic search and its application to feature-selection for deep relational machines", *International Conference on Artificial Neural Networks*, 2019.
   URL: https://doi.org/10.1007/978-3-030-30484-3_3

6. T. Dash, A. Srinivasan, L. Vig, O.I. Orhobor, R.D. King, "Large-scale assessment of deep relational machines", *International Conference on Inductive Logic Programming*, 2018.
   URL: https://doi.org/10.1007/978-3-319-99960-9_2
   (*Winner of the Best Student Paper Award)