Introduction to Deep Learning for Healthcare

Cao Xiao • Jimeng Sun

Introduction to Deep Learning for Healthcare

Springer

Cao Xiao
Seattle, WA, USA

Jimeng Sun
San Francisco, CA, USA

ISBN 978-3-030-82183-8 ISBN 978-3-030-82184-5 (eBook)
https://doi.org/10.1007/978-3-030-82184-5

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2021
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

> Life can only be understood backwards, but it must be lived forwards.
>
> *Søren Kierkegaard*

Deep learning models are multi-layer neural networks that have shown great success in diverse applications. This is a book describing deep learning models in the context of healthcare applications.

Story 1 When we took an artificial intelligence class many year ago, many topics were covered, including neural networks. The neural network model was presented as a supervised learning method. However, it was considered a practical failure compared to other more effective supervised learning methods such as decision trees and support vector machine. The common explanation about neural networks at the time involves two aspects: (1) Multi-layer neural networks can approximate any arbitrary functions and hence is a theoretically powerful model. (2) In practice, they don't work well due to the ineffective learning algorithm (i.e., backpropagation method). When we asked why backpropagation doesn't work well, a typical answer was about the accumulated errors across layers, which will eventually become too big to lead to an accurate model. Of course, the understanding of neural networks has evolved greatly in the past few years. When big labeled datasets and parallel computing infrastructure such as graphic processing units (GPU) finally become available, the power of deep neural networks will be unleashed. These days, deep learning models have become the most popular and standard machine learning models.

Story 2 When we first got into machine learning for healthcare many years ago, we spoke with a senior medical doctor about the potential impact of machine learning and artificial intelligence (AI) in medicine in the future. Specifically, we asked him about the possibility of creating AI algorithms to mimic the practice of real-world doctors. He was very pessimistic about the possibility because he believes

doctors largely depend on medical "intuition" to do their job, which is impossible to be learned by algorithms. Of course, now we know it is not only possible, but often AI algorithms can outperform human experts in various clinical pattern recognition tasks such as diagnosis. Even commercial medical devices have now become available (e.g., atrial fibrillation detection algorithm in Apple Watch). Many rely on deep learning models. Before we finished the book, we saw that doctor's profile on LinkedIn listed as an innovator in AI for healthcare.

Seattle, WA, USA Cao Xiao

Champaign, IL, USA Jimeng Sun

Contents

1 Introduction 1
- 1.1 Motivating Applications 2
  - 1.1.1 Diabetic Retinopathy Detection 3
  - 1.1.2 Early Detection of Heart Failure 4
  - 1.1.3 Sleep Analysis 4
  - 1.1.4 Treatment Recommendation 4
  - 1.1.5 Clinical Trial Matching 5
  - 1.1.6 Molecule Property Prediction and Generation 5
- 1.2 Who Should Read This Book? 5
- 1.3 Who Are the Authors? 6
- 1.4 Book Organization 6
- 1.5 Exercises 8

2 Health Data 9
- 2.1 The Growth of Electronic Health Records 9
- 2.2 Health Data 10
  - 2.2.1 The Life Cycle of Health Data 11
  - 2.2.2 Structured Health Data 13
  - 2.2.3 Unstructured Clinical Notes 15
  - 2.2.4 Continuous Signals 16
  - 2.2.5 Medical Imaging Data 17
  - 2.2.6 Biomedical Data for *In Silico* Drug Discovery 18
- 2.3 Health Data Standards 18
- 2.4 Exercises 21

3 Machine Learning Basics 23
- 3.1 Predictive Modeling Pipeline 23
- 3.2 Supervised Learning 25
  - 3.2.1 Logistic Regression 25
  - 3.2.2 Softmax Regression 27
  - 3.2.3 Gradient Descent 28
  - 3.2.4 Stochastic and Minibatch Gradient Descent 28

3.3 Unsupervised Learning 29
3.3.1 Principal Component Analysis 30
3.3.2 Clustering 31
3.4 Evaluation Metrics 31
3.4.1 Evaluation Metrics for Regression Tasks 31
3.4.2 Evaluation Metrics for Classification Tasks 33
3.4.3 Evaluation Metrics for Clustering Tasks 37
3.4.4 Evaluation Strategy 38
3.5 Exercises 39

4 Deep Neural Networks (DNN) 41
4.1 A Single Neuron 41
4.1.1 Activation Function 42
4.1.2 Loss Function 46
4.1.3 Train a Single Neuron 47
4.2 Multilayer Neural Network 49
4.2.1 Network Representation 49
4.2.2 Train a Multilayer Neural Network 50
4.2.3 Parameters and Hyper-Parameters 55
4.3 Case Study: Readmission Prediction from EHR Data with DNN 56
4.4 Case Study: DNN for Drug Property Prediction 57
4.5 Exercises 60

5 Embedding 63
5.1 Overview 63
5.2 Word2Vec 64
5.2.1 Idea and Formulation of Word2Vec 64
5.2.2 t-Distributed Stochastic Neighbor Embedding (t-SNE) 66
5.2.3 Healthcare Application of Word2Vec 68
5.3 Med2Vec: Two-Level Embedding for EHR 72
5.3.1 Med2Vec Method 72
5.4 MiME: Embed Internal Structure 75
5.4.1 Notations of MIME 75
5.4.2 Description of MIME 76
5.4.3 Experiment Results of MIME 78
5.5 Exercises 80

6 Convolutional Neural Networks (CNN) 83
6.1 CNN Intuition 83
6.2 Architecture of CNN 84
6.2.1 Convolution Layer: 1D 84
6.2.2 Convolution Layer: 2D 86
6.2.3 Pooling Layer 88
6.2.4 Fully Connected Layer 89

6.3 Backpropagation Algorithm in CNN* 89
6.3.1 Forward and Backward Computation for 1D Data 89
6.3.2 Special CNN Architectures 93
6.4 Case Study: Diabetic Retinopathy Detection 98
6.5 Case Study: Skin Cancer Detection 100
6.6 Case Study: Automated Surveillance of Cranial Images for Acute Neurologic Events 101
6.7 Case Study: Detection of Lymph Node Metastases from Pathology Images 103
6.8 Case Study: Cardiologist-Level Arrhythmia Detection and Classification in Ambulatory ECG 104
6.9 Case Study: COVID X-Ray Image Classification 105
6.10 Exercises 107

7 Recurrent Neural Networks (RNN) 111
7.1 RNN Fundamentals 111
7.2 Backpropagation Through Time (BPTT) Algorithm 115
7.2.1 Forward Pass 115
7.2.2 Backward Pass 115
7.3 RNN Variants 117
7.3.1 Long Short-Term Memory (LSTM) 117
7.3.2 Gated Recurrent Unit (GRU) 120
7.3.3 Bidirectional RNN 123
7.3.4 Encoder-Decoder Sequence-to-Sequence Models 124
7.4 Case Study: Early Detection of Heart Failure 125
7.5 Case Study: Sequential Clinical Event Prediction 127
7.6 Case Study: De-identification of Clinical Notes 129
7.7 Case Study: Learning to Prescribe Treatment Combination for Multimorbidity 130
7.8 Exercises 134

8 Autoencoders (AE) 137
8.1 Overview 137
8.2 Autoencoders 138
8.3 Sparse Autoencoders 139
8.4 Stacked Autoencoders 140
8.5 Denoising Autoencoders 141
8.6 Case Study: "Deep Patient" via Stacked Denoising Autoencoders 142
8.7 Case Study: Learning from Noisy, Sparse, and Irregular Clinical Data 143
8.8 Exercises 145

9 Attention Models 147
9.1 Overview 147
9.2 Attention Mechanism 147

9.3 Case Study: Attention Model over Longitudinal EHR 150
9.4 Case Study: Attention Model over a Medical Ontology 154
9.5 Case Study: ICD Classification from Clinical Notes 156
9.6 Case Study: Heart Disease Detection from Electrocardiography 158
9.7 Exercises 161

10 Graph Neural Networks 163
10.1 Overview 163
10.2 Notations and Tasks on Graphs 164
10.2.1 Notations and Operations 164
10.2.2 Tasks on Graphs 165
10.3 Graph Neural Networks 166
10.4 Graph Convolutional Networks 166
10.5 Message Passing Neural Network (MPNN) 167
10.6 Graph Attention Networks 168
10.7 Case Study: Neural Fingerprint in Drug Molecule Embedding with GCN 169
10.8 Case Study: *Decagon* Modeling Polypharmacy Side Effects with GCN 170
10.9 Case Study: Deep Learning Approach to Antibiotic Discovery 173
10.10 Case Study: *STAN* Spatio-Temporal Attention Network with GAT for Pandemic Prediction 175
10.11 Exercises 177

11 Memory Networks 181
11.1 Original Memory Networks 181
11.2 End-to-End Memory Networks 183
11.3 Self-Attention and Transformer 185
11.4 BERT: Pre-training of Deep Bidirectional Transformers 187
11.5 Case Study: Doctor2Vec—Doctor Recommendation for Clinical Trial Recruitment 188
11.6 Case Study: Medication Recommendation 191
11.7 Case Study: Pre-training of Graph Augmented Transformers for Medication Recommendation 196
11.8 Exercises 202

12 Generative Models 205
12.1 Generative Adversarial Networks (GAN) 205
12.1.1 The GAN Framework 206
12.1.2 The Loss Function of Discriminator 207
12.1.3 The Loss Function of Generator 207
12.1.4 Caveats of GAN 208

12.2 Variational Autoencoders (VAE) 208
12.2.1 VAE from Deep Learning Perspective 208
12.2.2 VAE from Probabilistic Model Perspective 210
12.2.3 Reparameterization Trick 213
12.3 Case Study: Generating Patient Records with GAN 214
12.4 Case Study: Molecule Generation Using VAE 217
12.5 Case Study: *MolGAN* an Implicit Generative Model for Small Molecular Graphs 219
12.6 Exercises 221

Bibliography 223