

М.Ю. Пришляк, С.А. Субботин, А.А. Олейник

Запорожский национальный технический университет, Запорожье

АНАЛИЗ ГЛУБОКИХ МОДЕЛЕЙ НЕЙРОННЫХ СЕТЕЙ НА БАЗЕ ОГРАНИЧЕННЫХ МАШИН БОЛЬЦМАНА

Проведен анализ моделей глубоких нейронных сетей и методов их обучения, в результате которого предложена классификация и дана сравнительная характеристика, а также выделены достоинства и недостатки архитектур глубоких нейросетей. Проведено экспериментальное исследование по обучению ограниченной и глубокой машин Больцмана, а также глубокой сети доверия для решения задачи классификации на примере задачи медицинского диагностирования рака груди с применением открытой библиотеки DeepLearning4J для построения нейромодели. Проведенные эксперименты позволили получить зависимости затрат памяти ЭВМ программных реализаций исследуемых методов от размерности выборок, а также сформулировать рекомендации по выбору архитектуры глубоких сетей при решении задач классификации.

Ключевые слова: глубокая нейронная сеть, машинное обучение, машина Больцмана, классификация.

Введение

Для автоматизации принятия решений в задачах технической и биомедицинской диагностики, распознавания графических изображений, текста, звука, сигналов разной природы необходимо решать задачу классификации данных по признакам, заключающуюся в определении принадлежности имеющих объектов, описываемых набором значений признаков, к определенным классам.

Для решения задач классификации могут применяться такие группы методов, как деревья решений [1], метод Байеса [2], искусственные нейронные сети (НС) [3], нейро-нечеткие сети [4] и другие. Среди данных методов наиболее перспективными являются НС, которые в отличие от метода Байеса и деревьев решений способны адаптироваться под природу входных данных и являются нелинейными моделями, что позволяет им в определенных рамках моделировать сложные распределения данных. В отличие от нейро-нечетких сетей, НС позволяют получать более компактные модели, обладающие высоким уровнем обобщения данных. Поэтому для задач построения классифицирующих моделей по наблюдениям целесообразно выбрать НС.

Для представления нейросетями сложных функций требуется большое число узлов и связей между ними, но увеличение их числа замедляет процесс обучения и может привести к переобучению сети [5]. Эту проблему можно решить применением глубоких НС. Эти сети образуются путем добавления в сеть скрытых слоев нейронов, повышающих вычислительную способность сети, но приносящих другие проблемы. К примеру, методы градиентного спуска при применении к глубоким НС могут находить неудачный локальный минимум

и оставаться в нем, уменьшая точность сети и её способности к обобщению. Если сеть инициализируется большими значениями весов, обучение обычно останавливается в локальной области параметров и метод обучения оказывается не способен изменить веса на достаточную величину для перехода в область параметров, представляющих оптимальное решение. Если же сеть инициализируется малыми весами, то они не позволяют значениям градиента функции ошибки, рассчитанным в выходном слое, распространяться в нижние слои и результирующая сеть работает, будто её нижние слои не обучены совсем [5]. Также, при использовании глубоких НС в практических задачах чаще всего заранее неизвестен размер сети, необходимой для достижения требуемой точности решения ею конкретной задачи, в результате чего можно получить сеть с чрезмерно большим числом параметров, что негативно влияет на скорость её обучения.

Целью работы является обзор и сравнение наиболее известных моделей глубоких НС, а также популярных методов обучения моделей на основе ОМБ – ГМБ и ГСД, на примере решения задачи классификации для определения их эффективности, достоинств и недостатков.

Постановка задачи

Пусть даны X – множество объектов и их признаков, Y – множество классов, к которым нужно отнести объекты множества X . Существует неизвестная зависимость $X \rightarrow Y$, реализации которой известны только на объектах конечной обучающей выборки $X^n = \{(x_1, y_1), \dots, (x_n, y_n)\}$.

Тогда задача построения распознающей модели в общем виде может быть сформулирована следую-

шим образом: необходимо определить функцию $F(X)$, которая будет способна определить принадлежность произвольного объекта $x \in X$ к соответствующему ему классу $y \in Y$ с ошибкой E , не превышающей заданный уровень ξ .

Применительно к НС задачу построения классификатора можно представить как задачу нахождения такой структуры нейромодели – зависимости $F(X, W)$ и значений её параметров W , при котором критерий качества модели – функция ошибки E не превысит заданное значение ξ .

Для случая НС-классификатора, модель зависимости в качестве параметров имеет набор значений весовых коэффициентов связей (весов) и пороговых смещений нейронов (порогов) W , а структурно состоит из входного слоя $V = \{v_i\}$, где v_i – узел видимого слоя, множества скрытых слоев сети

$H^k = \{h_j^k\}$, $k = 1, \dots, n$, где n – число скрытых слоев сети, и выходного слоя сети H^n . На входной слой сети подаются значения из множества X , последовательно по слоям передаются от входного слоя к выходному слою и далее сигналы с выходного слоя нейронов относятся ко множеству классов Y .

Классификация глубоких нейронных сетей

В настоящее время известно множество видов НС [6], среди которых глубокие сети можно рассматривать как класс парадигм, обладающих большим числом слоев. Известные глубокие НС предлагаются классифицировать следующим образом (рис. 1).

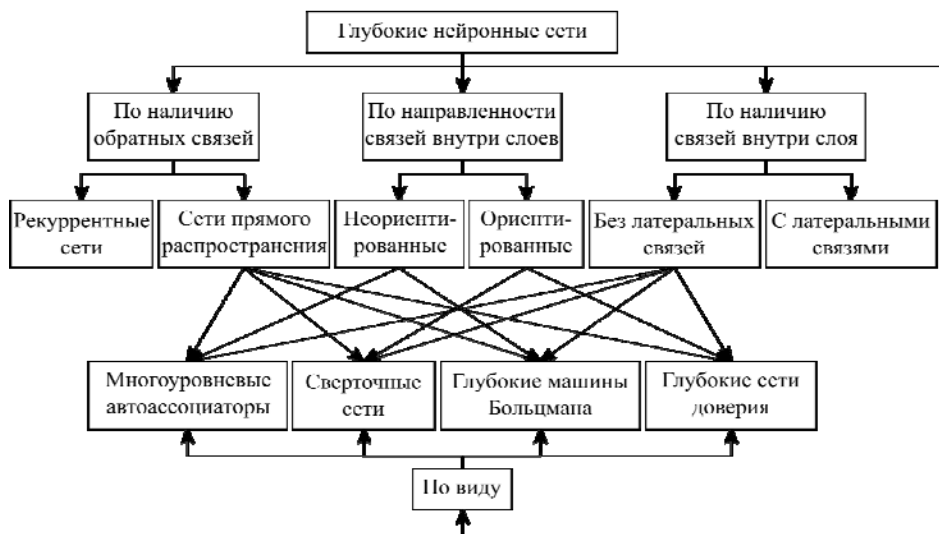


Рис. 1. Классификация и виды известных видов глубоких нейронных сетей

По наличию обратных связей в сети выделяют сети: прямого распространения (сигналы последовательно распространяются от входного слоя к выходному, отсутствуют обратные связи у нейронов) и рекуррентные (содержащие обратные связи).

По направленности связей внутри слоев сети существуют НС: ориентированные (существует заранее определенное направление у связей между нейронами), неориентированные (связи между нейронами не имеют направления и позволяют распространение значений в прямом и обратном направлениях).

По наличию связей между нейронами внутри слоя выделяют НС: с латеральными (боковыми) связями между нейронами внутри одного слоя и без латеральных связей.

Рассмотрим наиболее широко используемые на практике архитектуры глубоких НС.

Ограниченные машины Больцмана

ОМБ – это НС, состоящая из двух слоев – видимого слоя v и скрытого h , у которых отсутствуют

связи внутри слоя (потому сеть называется "ограниченной"), и состоящих из бинарных узлов с сигмоидальной функцией активации [7].

Схема простейшей ОМБ показана на рис. 2.

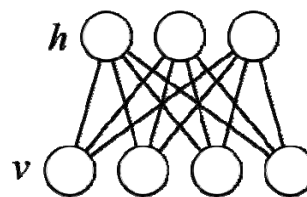


Рис. 2. Схема ОМБ

Поскольку ОМБ не содержит связей внутри слоя, то активация каждого узла внутри слоя взаимно независима и зависит лишь от значений, полученных с другого слоя. В общем виде вероятность активации узла в скрытом слое можно представить в виде уравнения (1):

$$P(h_j = 1 | v) = \sigma \left(b_j + \sum_{i=1}^m W_{i,j} v_i \right); \quad (1)$$

а узла в видимом слое – уравнением (2):

$$P(v_i = 1 | h) = \sigma \left(a_i + \sum_{j=1}^n W_{i,j} h_j \right); \quad (2)$$

где m , n – размеры видимого и скрытого слоев соответственно; a , b – векторы смещений видимого и скрытого слоев; v_i , h_j – узлы видимого и скрытого слоев; W – матрица весов связей; а σ – логистическая (сигмоидальная) функция.

Базовыми методами для обучения ОМБ являются выборка по Гиббсу [8] и контрастная дивергенция [8].

Метод выборки по Гиббсу является одним из методов выборки Монте-Карло по схеме марковских цепей (MCMC, Markov-Chain Monte-Carlo). Метод состоит из двух последовательно выполняемых этапов – выборки значений скрытых узлов сети следуя уравнению (1), затем выборки значений видимых узлов по уравнению (2). Эти два этапа представляют собой один шаг выборки. Выборка по Гиббсу повторяется необходимое количество шагов, либо до наступления сходимости марковской цепи, после чего выполняется обновление параметров сети.

Контрастная дивергенция является частным случаем выборки по Гиббсу, рассчитанным на ускорение процесса обучения за счет использования аппроксимации стохастического градиента вместо его прямого вычисления [8]. Этот метод не требует достижения сходимости марковской цепи и эффективен даже при выполнении одного шага выборки по Гиббсу [9], потому является достаточно быстрым для обучения ОМБ.

Достоинствами ОМБ являются: генеративная природа сети – может применяться для моделирования неизвестного заранее распределения; после обучения сеть способна работать с неполными данными (и дополнять их при необходимости), а также ОМБ можно применять для обучения других, более сложных моделей.

Недостатками модели является длительность времени обучения – сходимость выборки по Гиббсу является медленным процессом.

Глубокие машины Больцмана

ГМБ – сеть, состоящая из слоя видимых и нескольких слоев скрытых узлов, причем связи присутствуют лишь между узлами из смежных слоев. ГМБ преимущественно обучаются без учителя на большом количестве данных, а обучение с учителем применяется для подстройки сети под конкретную задачу. Схема классической ГМБ показана на рис. 3, а.

ГМБ Гаусса-Бернулли (ГТМБ) – ГМБ, у которой в видимом слое применяются Гауссовы узлы. Без тщательного выбора параметров обучения эта модель может столкнуться с проблемами на этапе

подстройки [10]. Схема этой сети идентична классической ГМБ, за исключением типа применяемых в видимом слое узлов.

Многопрогнозная ГМБ (МПГМБ) образуется при применении многопрогнозного обучения к ГМБ. Полученная модель может рассматриваться как единая вероятностная модель, обученная с помощью вариационного критерия, либо как семейство рекуррентных сетей, которые решают связанные задачи вывода. Данная НС не нуждается в жадном послонном обучении, а качество решения ею определенной задачи может быть определено заранее. Также МПГМБ не требует слой классификаторов для решения задачи логического вывода [11].

Разреженная групповая ГМБ (РГТМБ) отличается от других разновидностей ГМБ тем, что выполняет прореживание входных данных в пределах небольших групп с использованием регуляризатора. Внутри групп она также стимулирует конкуренцию между скрытыми узлами. Разработка этой модели была направлена на ускорение процесса обучения и повышение качества классификации [12]. Схема этой сети визуально похожа на классическую ГМБ, отличием является применение регуляризатора в пределах каждой ОМБ в процессе обучения.

Все рассмотренные виды ГМБ могут применяться для решения задачи классификации с различной точностью. ГТМБ эффективна в работе с вещественными данными без предварительной нормализации значений до бинарных. МПГМБ полезна для решения задачи классификации с неполными данными и задач прогнозирования среднего поля. РГТМБ в основном рассчитаны только на решение задачи классификации, но с большей точностью, чем классические ГМБ.

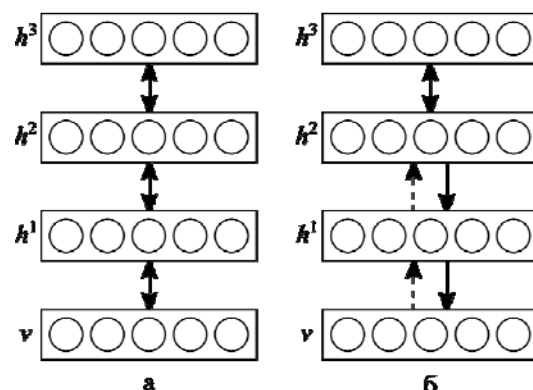


Рис. 3. Схематическое изображение НС:
а – ГМБ; б – ГСД

Для обучения ГМБ существуют различные методы.

Жадное послонное обучение [13] является наиболее распространенным методом для обучения ГМБ. В нем каждая ОМБ обучается отдельно от других, используя при этом значения, полученные с

заранее обученной предыдущей ОМБ. Его достоинством является быстрая инициализация весов сети [13], но увеличение числа слоев в модели может усложнять процесс обучения.

Дифференциальная подстройка ГМБ (discriminative fine-tuning) [13] позволяет улучшить качество решения сетью задачи классификации, заменив стохастическую активность бинарных признаков в слоях на вещественные вероятности. Метод полезен при обучении без учителя и требует небольшое количество помеченных данных для подстройки результирующей сети [13].

Многопрогнозовое обучение (multi-prediction training) [11] использует уравнения среднего поля в ГМБ, которая затем применяется для обучения рекуррентных нейронных сетей решению задач логического вывода. Одним из достоинств этого метода обучения является отсутствие состояния при обучении – метод не ссылается на данные с предыдущих итераций обучения. Это означает, что точность градиента многопрогнозового обучения не зависит от таких свойств метода, как скорость обучения и т.п. Другим достоинством является то, что метод обучения применим для разных типов узлов – бинарных, Гауссовых и Softmax. К недостаткам можно отнести вычислительную сложность – для достижения сходимости уравнений приближения среднего поля (mean-field inference), необходимо выполнять расчеты несколько раз [11].

Адаптивная искусственная закалка (adaptive simulated tempering) [14] использует метод Вангала-Ландау для равномерного распределения значений в поле энергии. Достоинством метода является возможность моделирования распределения многомерных и структурированных данных – он позволяет сети покидать область одного локального минимума (в котором могут останавливаться другие методы обучения) для проведения сетью равных промежутков времени в других локальных минимумах. Недостатком метода является его скорость – он примерно в два раза медленнее выборки по Гиббсу [14].

Центрирование (the centering trick) [15] предлагает замену функции энергии на функцию центрированных состояний. Применение такой функции способствует уменьшению структурной и аппроксимационной нестабильностей процесса обучения, что является основным достоинством метода и позволяет создавать более устойчивые к шуму модели. Недостатком является возможность применения этого метода для обучения небольших моделей. В случае моделей размерностью в тысячи узлов этот метод применим только с использованием регуляризатора – послонного обучения либо прореживания связей [15].

Среди рассмотренных моделей НС классические ГМБ являются универсальной сетью для реше-

ния задач классификации и генерации данных. ГМБ эффективны для запоминания и реконструкции изображений (человеческих лиц, трехмерных объектов на изображении), но являются менее эффективными в решении задачи классификации, чем остальные сети. МПГМБ используют в основном для решения задач классификации данных в общем, как с полными, так и с недостающими значениями. РГГМБ применимы для классификации изображений и для предобучения других видов глубоких сетей решению задачи классификации.

Для решения задач диагностирования по признакам, представленным в виде числовых массивов данных, целесообразно применять классические либо многопрогнозовые ГМБ. Среди рассмотренных моделей ГМБ для решения задач, где велико число признаков, целесообразно применять РГГМБ за счет их способности работать с избыточными данными и наличия функции-регуляризатора.

Глубокие сети доверия

ГСД – это глубокая, стохастически обучаемая модель, верхние два слоя которой формируют модель ОМБ и имеют двунаправленные связи, а остальные формируют направленную глубокую сеть [16]. Такие сети могут использоваться в качестве дифференцирующих и генеративных моделей, а их основное назначение – выделение признаков и сжатие данных. Схема классической ГСД показана на рис. 3, б. К разновидностям ГСД относят следующие сети.

Разреженная ГСД (РГСД, sparse DBN) – это глубокая сеть, состоящая из разреженных ОМБ в качестве слоев. В разреженных ОМБ обычно применяется функция-регуляризатор, которая поддерживает активации узлов на уровне, близком к нулю [17].

Сверточная ГСД (convolutional DBN) состоит из сверточных ОМБ и применяется в основном для работы с полноразмерными изображениями и аудиоданными. Они применяют подход вероятностного объединения максимумов (probabilistic max-pooling) и связанные веса, что позволяет узлам верхних слоев сети выделять общие признаки в различных частях изображения. Сеть обычно обучается жадным, послонным методом [18].

Стимулируемая ГСД (boosted DBN) состоит из нескольких классических ГСД, каждая из которых обучается иерархическому представлению признаков на одинаковом наборе частей изображений. Эти модели связаны через стимулирующий классификатор и подстраиваются вместе с использованием единой целевой функции так, что признаки, извлеченные из разных местоположений во входных данных, выбираются и усиливаются вместе в зависимости от их относительной важности [19].

Рассмотрим известные методы обучения ГСД.

Жадное послонное обучение [16] – метод, в котором слои НС обучаются последовательно, друг за другом. После обучения одного слоя его веса фиксируются, а значения, получаемые с него, используются для обучения следующего слоя. Метод быстрый и способен обучать генеративные модели, как без учителя, так и с применением маркированных данных, получая в результате модель, генерирующую данные и соответствующие метки, но этот метод задает значения только генеративным весам сети [16].

Подстройка сети [16] – метод, обычно применяемый после жадного обучения сети для корректировки распознающих весов сети. В нем используется подход "пробуждения-засыпания", состоящий из двух последовательно применяемых фаз, корректирующих генеративные и распознающие веса сети соответственно. Метод помогает подстроить сеть для решения конкретной задачи. Недостатком метода является его скорость – в нем применяется выборка по Гиббсу для верхнего слоя ГМБ [16].

Обучение с регуляризацией разреженности [18] (training via sparsity regularization) использует регуляризацию целевой функции для приближения среднего значения активаций всех скрытых узлов к некоторой малой константе. Метод способен выполнять обучение сети для работы с избыточным количеством входных данных (например, с большими изображениями) и предотвращать такие модели от обучения тривиальным решениям. Применение связанных весов и подхода вероятностного объединения максимумов ускоряет процесс обучения [18]. Однако этот метод работает с модифицированной моделью ГСД, которая содержит неориентированные связи между всеми слоями.

Среди рассмотренных архитектур НС классические ГСД в основном используют для решения задач классификации изображений, но они также применимы и для классификации массивов данных, представленных в числовом виде.

РГСД применимы для решения таких задач, как классификация изображений, классификация коротких текстов на естественных языках и распознавание речи.

Сверточные ГСД применимы для большинства задач визуального распознавания – их возможно обучать классификации изображений как с учителем, так и без него, а также применять для решения задач выделения иерархических представлений из изображений, и распознавать изображенные на них объекты.

Стимулируемые ГСД в основном эффективны для решения задач распознавания выражений человеческих лиц, но не ограничиваются этой задачей и

могут применяться для решения других задач классификации изображений.

В качестве универсальной сети для решения задач диагностирования по признакам, не имеющим представления в виде изображений и звука, целесообразно использовать классические ГСД, поскольку они представляют собой универсальную сеть для решения задач классификации и генерации. Среди архитектур данного типа в задачах, где имеется большое число признаков, целесообразно использовать разреженные ГСД, поскольку они способны выполнять регуляризацию процесса обучения и "прореживание" входных данных для предотвращения сети от переобучения.

В табл. 1 представлена сравнительная характеристика ГМБ и ГСД.

Таблица 1
Сравнение ГМБ и ГСД

Характеристика	ГМБ	ГСД
Направленность связей	Неориентированные	Ориентированные
Наличие обратных связей	Нет	В гибридных моделях
Наличие связей внутри слоя	Нет	Нет
Необходимость предобучения	Рекомендуется	Рекомендуется
Генеративная модель	Да	Да
Дифференцирующая модель	Да	Да
Шумоподавление	Нет	Нет
Работа с избыточными данными	РГМБ	РГСД, сверточные ГСД
Работа с изображениями	Да	Да

Жадные обучающие процедуры для ГМБ и ГСД состоят из двух фаз – предобучения и корректировки.

На фазе предобучения сети представляются в виде набора ограниченных машин Больцмана (ОМБ), которые обучаются последовательно. Этот процесс происходит следующим образом. Сначала нижние два слоя (v и h_1) представляются в виде ОМБ, выполняется её обучение с использованием входных данных X , в результате чего получаем матрицу весов W . Далее значения этих весов закрепляются и выполняется обучение следующих двух слоев как ОМБ с использованием данных, получаемых с предыдущего обученного слоя при подаче входных значений X на видимый слой. Этот процесс повторяется для всех пар слоев сети до тех пор, пока не будет достигнут последний.

На фазе корективної виконується підстройка мережі методом навчання з учителем.

Розглянуті глибокі моделі НС застосовуються для вирішення різних завдань: класифікація по числовим даним, класифікація зображень і об'єктів на них, відновлення відсутніх (пошкоджених) частин зображень, розпізнавання мови, обробка текстів на природних мовах. В частині, РГМБ, РГСД і сверточні ГСД ефективні для вирішення завдання класифікації і роботи з великими зображеннями, стимульовані ГСД підходять для класифікації зображень (розпізнавання виразів людських обличчя), ГМБ застосовуються для запам'ятовування і реконструкції зображень, а МПГМБ здатні вирішувати завдання класифікації з неповними даними. Для інших завдань цілком доцільно застосовувати класичні ГМБ і ГСД, або інші спеціалізовані моделі, не розглянуті в цій роботі.

Серед розглянутих методів навчання глибоких НС найбільш універсальними і просто реалізуєми на практиці є жадні методи навчання, тому вони цікаві для подальшого вивчення на практичних завданнях.

Проблеми існуючих методів навчання глибоких НС полягають в тому, що універсальні методи (такі як жадне навчання) можуть не надавати достатньої точності рішення конкретної задачі, а узкоспеціалізовані методи навчання хоч і здатні вирішувати певні завдання з достатньою точністю, але їх важко застосовувати для інших завдань через введені обмеження.

Експерименти

Для дослідження властивостей розглянутих методів були проведені експерименти по вирішенню практичних завдань класифікації. Обчислювальні експерименти виконувалися з використанням відкритої бібліотеки DeepLearning4J [20], для мови програмування Java. Бібліотека здатна проводити навчання нейронних мереж довільної архітектури, а сам процес навчання можна виконувати паралельно в кластері або на графічному адаптері.

В якості вихідних даних був взят набір даних з публічного репозитора для діагностики раку грудей [21]. Цей набір складається з 32 ознак і 569 зразків. Результатом рішення задачі є постановка діагнозу – наявність або відсутність раку.

В експериментах досліджувалися наступні архітектури НС: ОМБ, ГМБ і ГСД. Для кожної мережі були вибрані наступні конфігурації шарів: для ОМБ – один шар з 30 входами і 2 виходами; для ГМБ – 4 шари (30 нейронів у входному шарі, в наступних, відповідно – 5, 4, 3 і 2 виход-

них нейрона); для ГСД – 4 шари (конфігурація, аналогічна ГМБ).

Для отримання достатньої кількості даних для подальшого вивчення, для навчання був встановлений межа в 1500 епох для всіх НС, в результаті чого вибиралася найкраща епоха (перша епоха, показавши найкращий результат). Вхідні дані для всіх моделей розділялися на дві вибірки – навчальну (40%) і тестову (60%). Над даними виконувалася нормалізація – значення ознак приводилися до такого виду, щоб середнє значення було рівно нулю, а середньквдратичне відхилення було рівно одиниці. Навчання всіх мереж виконувалося жадним методом.

В кінці кожної епохи на етапі навчання виконувалося вимірювання часу і використання ОП комп'ютера, а також розраховувалося оцінка ймовірності прийняття правильного рішення для обох вибірок (навчальної і тестової).

Експерименти проводилися на обладнанні з наступними характеристиками: процесор Intel Core i5-4570K, 16 Гб оперативної пам'яті DDR3-1600, графічний адаптер GeForce GTX 1060 6GB. Графічний адаптер в тестах не використовувався, тому що його застосування приводило тільки до уповільнення процесу навчання.

Результати експериментів

По результатам проведених експериментів були побудовані графіки залежності середньквдратичної помилки від номера епохи на навчальних даних (рис. 4), графіки оцінки ймовірності прийняття правильного рішення НС в процесі навчання по навчальній (рис. 5) і тестовій вибіркам (рис. 6), а також графік залежності використання оперативної пам'яті (ОП) в мегабайтах бібліотекою в процесі навчання (рис. 7). Чисельні результати експериментів наведені в табл. 2.

Таблиця 2
Результати порівняння моделей НС при вирішенні задачі діагностики

Модел	Час навчання до заданої епохи, сек.	Номер найкращої епохи	Час навчання до найкращої епохи, сек.	Оцінка ймовірності прийняття правильного рішення, %
ОМБ	68,83	451	22,5	96,49
ГМБ	98,03	1235	75,49	97,08
ГСД	72,93	1499	72,93	97,37

Обговорення

По графіку зміни середньквдратичної помилки НС в часі навчання (рис. 4) видно, що ОМБ навчається досить швидко (в епоху 451 у неї було найменше значення помилки).

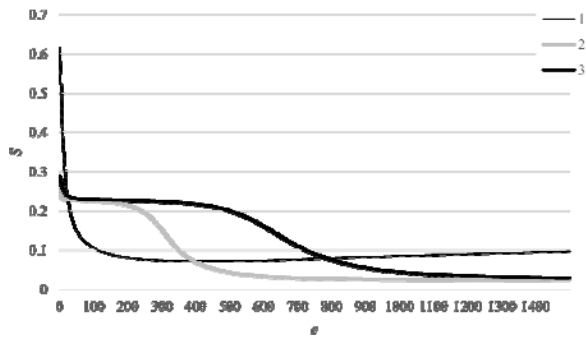


Рис. 4. График зависимости среднеквадратичной ошибки S от номера эпохи e :
1 – ОМБ, 2 – ГМБ, 3 – ГСД

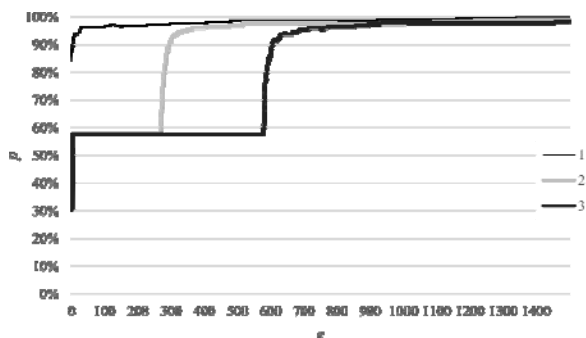


Рис. 5. График зависимости оценки вероятности принятия правильного решения p на обучающей выборке от номера эпохи e :
1 – ОМБ, 2 – ГМБ, 3 – ГСД

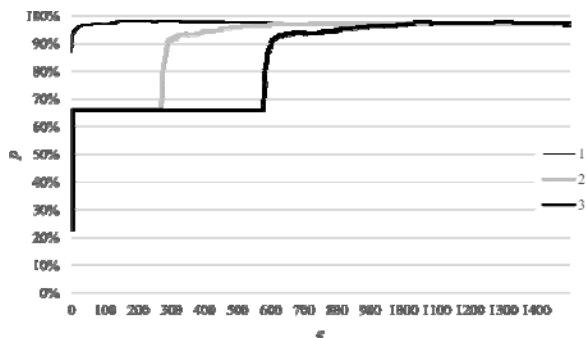


Рис. 6. График зависимости оценки вероятности принятия правильного решения p на тестовой выборке от номера эпохи e :
1 – ОМБ, 2 – ГМБ, 3 – ГСД

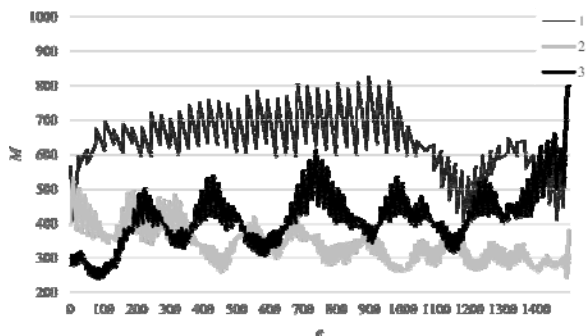


Рис. 7. График использования оперативной памяти M от номера эпохи e :
1 – ОМБ, 2 – ГМБ, 3 – ГСД

Однако отсутствие скрытых слоев сказывается на способности сети обучаться в дальнейшем, и её ошибка начинает расти. ГМБ и ГСД, в отличие от ОМБ, смогли минимизировать ошибку до меньшего значения, но им обоим на это потребовалось больше времени (1235 и 1499 эпох, соответственно).

Графики оценки вероятности принятия правильных решений (рис. 5–6) показывают, что ОМБ еще на первых итерациях процесса обучения достигла точности более 90%. ГСД проигрывает ГМБ по скорости обучения, но в результате все сети достигли высокой точности к 1000 эпохам и далее. Также по этим графикам заметно, что у каждой из моделей на определенной итерации обучения резко повышалась вероятность принятия правильных решений, причем у ГМБ этот момент наступил раньше, чем у ГСД.

На графике использования ОП (рис. 7) видно, что ОМБ, несмотря на свою простоту, при обучении требовала больших затрат памяти, по сравнению с другими сетями. ГМБ и ГСД использовали примерно одинаковый объем памяти в начале процесса обучения, но к середине процесса обучения и далее использование оперативной памяти ГСД возросло. Волнообразный вид этого графика обуславливается особенностью программной реализации на основе Java-машины – наличием сборщика мусора, который запускается автоматически для очистки неиспользуемых объектов в памяти процесса.

Из таблицы 2 видно, что до максимальной 1500 эпохи быстрее всех обучалась ОМБ, несколько медленнее обучались ГСД (на 5,9% дольше), и ГМБ (на 42,4% дольше ОМБ). До эпохи с наилучшей точностью быстрее всех также обучалась ОМБ, следом за ней идет ГСД (в 3,36 раза дольше) и ГМБ (в 3,24 раза дольше ОМБ). Примечательно, что разница во времени обучения до эпохи с наилучшим результатом у ГМБ и ГСД минимальна: ГМБ всего на 3,5% медленнее.

По вероятности принятия правильных решений лучше всех себя показала ГСД. Точность ГМБ оказалась на 0,29% ниже, а ОМБ – на 0,88% ниже ГСД.

Проведенные эксперименты показали, что среди трех рассмотренных моделей сетей наиболее быстро обучаемой является модель ОМБ, наиболее точной является ГСД. По графикам видно, что ОМБ эффективно применять в тех случаях, когда важна скорость обучения, но не точность. Среди рассмотренных глубоких моделей более эффективной как по скорости обучения, так и по оценке вероятности принятия правильного решения оказалась ГСД.

Также при выборе модели НС для решения конкретной задачи стоит учитывать сложность задачи. Однослойные НС (в том числе ОМБ) ограничены в возможности выделения сложных зависимостей во входных данных. Глубокие НС требуют

больше времени для обучения, но при этом работают с более высокой точностью.

Исследования показали, что для решения задач диагностирования по признакам, имеющим представления в виде массивов числовых значений различных измерений, эффективно применять классические ГСД. Среди рассмотренных сетей они показали лучшую точность классификации и оптимальное время обучения.

Выводы

В работе рассмотрена задача построения распознающих моделей на основе глубоких нейронных сетей. Проанализированы наиболее известные модели и методы обучения глубоких НС – ГМБ и ГСД.

Практическая ценность полученных результатов состоит в том, что проведено экспериментальное исследование рассмотренных методов построения глубоких НС. В результате экспериментов была определена их эффективность при обработке данных для медицинского диагностирования по признакам, изучены зависимости затрат памяти программных реализаций исследуемых методов в зависимости от характеристик входных данных.

Эксперименты по решению задачи классификации показали, что ОМБ эффективны в случае, когда точность менее важна, чем скорость обучения. Глубокие модели напротив, позволяют обучить более точные модели, причем ГМБ использовала наименьшие затраты оперативной памяти в процессе обучения, а ГСД обучалась быстрее ГМБ.

Описанные достоинства и недостатки известных моделей ГМБ и ГСД и методов их обучения отражают необходимость их улучшения для решения более сложных задач и получения лучших результатов при решении различных задач в сфере искусственного интеллекта.

Перспективы дальнейших исследований состоят в детальном изучении глубоких моделей НС при решении задач, отличных от технического и медицинского диагностирования. Также представляет интерес рассмотреть особенности применения классических и узконаправленных методов обучения к данным различной размерности и вида (например, к изображениям и звуковой информации).

Список литературы

1. Strobl C. An Introduction to Recursive Partitioning: Rationale, Application and Characteristics of Classification and Regression Trees, Bagging and Random Forests / C. Strobl, J. Malley, G. Tutz // *Psychological Methods*. – 2009. – Vol. 14, № 4. – P. 323-348. doi: 10.1037/a0016973.
2. An empirical study of the naive Bayes classifier: report / Thomas J. Watson Research Center; I. Rish. – New York, 2001. – 6 p. doi: 10.1.1.330.2788.
3. Zhang G.P. Neural networks for classification: a survey / G.P. Zhang // *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*. – 2000. – Vol. 30, № 4. – P. 451-462. doi: 10.1109/5326.897072.
4. Ghosh A. A novel approach to neuro-fuzzy classification / A. Ghosh, B. Uma Shankar, Saroj K. Meher // *Neural Networks*. – 2009. – Vol. 22, № 1. – P. 100-109. doi: 10.1016/j.neunet.2008.09.011.
5. Kukačka M. Overview of Deep Neural Networks / M. Kukačka // 21st Annual Conference of Doctoral Students – WDS 2012, Prague, 29 May–1 June 2012 : proceedings. Prague: MATFYZPRESS, 2012. – Part 1. – P. 100-105.
6. Горбачевская Е.Н. Классификация нейронных сетей / Е.Н. Горбачевская // *Вестник ВУиТ*. – 2012. – №2 (19). – С. 128-134.
7. Hinton G.E. Reducing the dimensionality of data with neural networks / G.E. Hinton, R.R. Salakhutdinov // *Science*. – 2006. – Vol. 313, № 5786. – P. 504-507. doi: 10.1126/science.1127647.
8. Cho K. Improved learning algorithms for restricted Boltzmann machines : thesis ... M.Sc. : Computer Science and Engineering / Cho Kyung Hyun. – Espoo: Aalto University, 2011. – 84 p. doi: 10.1007/978-3-642-21735-7_2.
9. Senoussaoui M. First attempt of boltzmann machines for speaker verification / M. Senoussaoui // *Odyssey 2012: The Speaker and Language Recognition Workshop, Singapore, 25–28 June 2012: proceedings*. – Singapore – ISCA Archive, 2012. – P. 117-121.
10. Cho K.H. Gaussian-Bernoulli deep Boltzmann machine / K.H. Cho, T. Raiko, A. Ilin // 2013 International Joint Conference on Neural Networks, Dallas, 04–09 August 2013: proceedings. – Los Alamitos: IEEE, 2013. – P. 1-7. doi: 10.1109/IJCNN.2013.6706831.
11. Goodfellow I. Multi-prediction deep Boltzmann machines / I. Goodfellow, M. Mirza, A. Courville et al. // *Neural Information Processing Systems 26, Lake Tahoe, 05-10 December 2013: proceedings*. – 2013. – P. 548-556. doi: 10.1109/MVA.2015.7153145.
12. Luo H. Sparse Group Restricted Boltzmann Machines / H. Luo, R. Shen, C. Niu et al. // 25th Conference on Artificial Intelligence, San Francisco, 7–11 August 2011: proceedings. – San Francisco: AAAI Press, 2011. – P. 429-434.
13. Salakhutdinov R. Deep Boltzmann Machines / R. Salakhutdinov, G. Hinton // 12th International Conference on Artificial Intelligence and Statistics, Hilton Clearwater Beach Resort, 16–18 April 2009: proceedings. – Clearwater Beach: PLMR, 2009. – Vol. 5. – P. 448-455.
14. Salakhutdinov R. Learning deep Boltzmann machines using adaptive MCMC / R. Salakhutdinov // 27th International Conference on Machine Learning, Haifa, 21–24 June 2010: proceedings. – Haifa: IMLS – 2010. – P. 943-950.

15. Montavon G. Deep Boltzmann machines and the centering trick / G. Montavon, M. Klaus-Robert // *Neural Networks: Tricks of the Trade: Second Edition* – Berlin : Springer, 2012. – Section 25. – 769 p. doi: 10.1007/978-3-642-35289-8_33
16. Hinton G.E. A fast learning algorithm for deep belief nets / G.E. Hinton, S. Osindero, Y. Teh // *Neural computation*. – 2006. – Vol. 18, № 7. – P. 1527-1554. doi: 10.1162/neco.2006.18.7.1527.
17. Keyvanrad M.A. Effective sparsity control in deep belief networks using normal regularization term / M.A. Keyvanrad, M.M. Homayounpour // *Knowledge and Information Systems*. – 2017. – Vol. 53, № 2. – P. 533-550. doi: 10.1007/s10115-017-1049-x.
18. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations / H. Lee, R. Grosse, R. Ranganath et al.] // *26th International Conference on Machine Learning, Montreal, 14–18 June 2009: proceedings*. – Montreal: ACM, 2009. – P. 609-616. doi: 10.1145/1553374.1553453.
19. Facial expression recognition via a boosted deep belief network / [P. Liu and S. Han and Z. Meng et al.] // *27th IEEE Conference on Computer Vision and Pattern Recognition, Columbus, 23–28 June 2014 : proceedings*. – Columbus : IEEE, 2014. – P. 1805–1812. doi: 10.1109/CVPR.2014.233
20. DeepLearning4j: Open-source, Distributed Deep Learning for the JVM [Електронний ресурс]. – Режим доступа к ресурсу: <https://deeplearning4j.org>.
21. Breast Cancer Wisconsin (Diagnostic) Data Set [Електронний ресурс]. – Режим доступа к ресурсу: [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)).

References

1. Strobl, C., Malley, J. and Tutz, G. (2009), An Introduction to Recursive Partitioning: Rationale, Application and Characteristics of Classification and Regression Trees, Bagging and Random Forests, *Psychological Methods*, Vol. 14, No. 4, pp. 323-348. <https://doi.org/10.1037/a0016973>.
2. Rish, I. (2001), *An empirical study of the naive Bayes classifier: report*, Thomas J. Watson Research Center, New York, 6 p. <https://doi.org/10.1.1.330.2788>.
3. Zhang, G.P. (2000), Neural networks for classification: a survey, *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, Vol. 30, No. 4, pp. 451-462. <https://doi.org/10.1109/5326.897072/>
4. Ghosh, A. (2009), A novel approach to neuro-fuzzy classification, *Neural Networks*, Vol. 22, No. 1, pp. 100-109. <https://doi.org/10.1016/j.neunet.2008.09.011/>
5. Kukačka, M. (2012), Overview of Deep Neural Networks, *21st Annual Conference of Doctoral Students – WDS 2012*, May 29–June 1, Prague, pp. 100-105.
6. Gorbachevskaja, E.N. (2012), "Klassifikacija nejronnyh setej" [Neural networks classification], *Vestnik VUiT*, No. 2, pp. 128-134.
7. Hinton, G.E. (2006), Reducing the dimensionality of data with neural networks, *Neural Networks*, Vol. 22, No. 1, pp. 100-109. <https://doi.org/10.1126/science.1127647>.
8. Cho, K. (2011), *Improved learning algorithms for restricted Boltzmann machines: thesis ... M.Sc.: Computer Science and Engineering*, Espoo: Aalto University, 84 p. https://doi.org/10.1007/978-3-642-21735-7_2.
9. Senoussaoui, M. (2012), First attempt of Boltzmann machines for speaker verification, *Odyssey 2012: The Speaker and Language Recognition Workshop*, June 25–28, Singapore, pp. 117-121.
10. Cho, K.H., Raiko, T. and Ilin, A. (2013), Gaussian-Bernoulli deep Boltzmann Machine, *2013 International Joint Conference on Neural Networks*, August 4–9, Dallas, pp. 1-7. <https://doi.org/10.1109/IJCNN.2013.6706831>.
11. Goodfellow, I., Mirza, M. and Courville, A. (2013), Multi-prediction deep Boltzmann machines, *Neural Information Processing Systems 26*, December 5–10, Lake Tahoe, pp. 548-556. <https://doi.org/10.1109/MVA.2015.7153145>.
12. Luo, H., Shen, R. and Niu, C. (2011), Sparse Group Restricted Boltzmann Machines, *25th Conference on Artificial Intelligence*, August 7–11, Lake Tahoe, pp. 548-556.
13. Salakhutdinov, R. and Hinton, G. (2009), Deep Boltzmann Machines, *12th International Conference on Artificial Intelligence and Statistics*, April 16–18, Clearwater Beach, pp. 448-455.
14. Salakhutdinov, R. (2010), Learning deep Boltzmann machines using adaptive MCMC, *27th International Conference on Machine Learning*, June 21–24, Haifa, pp. 943-950.
15. Montavon, G. and Klaus-Robert, M. (2012), *Neural Networks: Tricks of the Trade: Second Edition*, Springer, Berlin, 769 p. https://doi.org/10.1007/978-3-642-35289-8_33.
16. Hinton, G.E., Osindero, S. and Teh Y. (2006), A fast learning algorithm for deep belief nets, *Neural computation*, Vol. 18, No. 7, pp. 1527-1554. <https://doi.org/10.1162/neco.2006.18.7.1527>.
17. Keyvanrad, M.A. and Homayounpour, M.M. (2017), Effective sparsity control in deep belief networks using normal regularization term, *Knowledge and Information Systems*, Vol. 53, No. 2, pp. 533-550. <https://doi.org/10.1007/s10115-017-1049-x>.
18. Lee, H., Grosse, R. and Ranganath, R. (2010), Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations, *26th International Conference on Machine Learning*, 14–18 June, 2009, Montreal, pp. 609-616. <https://doi.org/10.1145/1553374.1553453>.
19. Liu, P., Han, S. and Meng, Z. (2014), Facial expression recognition via a boosted deep belief network, *27th IEEE Conference on Computer Vision and Pattern Recognition*, 23–28 June 2014, Columbus, pp. 1805-1812. <https://doi.org/10.1109/CVPR.2014.233>.
20. DeepLearning4j: Open-source, Distributed Deep Learning for the JVM, <https://deeplearning4j.org>.
21. Breast Cancer Wisconsin (Diagnostic) Data Set, [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)).

Відомості про авторів:

Пришляк Михайло Юрійович

аспірант Запорізького національного технічного університету, Запоріжжя, Україна
<https://orcid.org/0000-0002-7158-0394>
e-mail: pryshliak.mykhailo@gmail.com

Субботін Сергій Олександрович

доктор технічних наук професор завідувач кафедри Запорізького національного технічного університету, Запоріжжя, Україна
<https://orcid.org/0000-0001-5814-8268>
e-mail: subbotin.csit@gmail.com

Олійник Андрій Олександрович

кандидат технічних наук доцент кафедри Запорізького національного технічного університету, Запоріжжя, Україна
<https://orcid.org/0000-0002-6740-6078>
e-mail: olejnikaa@gmail.com

Information about the authors:

Mykhailo Pryshliak

Postgraduate Student of Zaporizhzhia National Technical University, Zaporizhzhia, Ukraine
<https://orcid.org/0000-0002-7158-0394>
e-mail: pryshliak.mykhailo@gmail.com

Sergey Subbotin

Doctor of Technical Sciences Professor Head of Department of Zaporizhzhia National Technical University, Zaporizhzhia, Ukraine
<https://orcid.org/0000-0001-5814-8268>
e-mail: subbotin.csit@gmail.com

Andrii Oliinyk

Candidate of Sciences Associate Professor of Department of Zaporizhzhia National Technical University, Zaporizhzhia, Ukraine
<https://orcid.org/0000-0002-6740-6078>
e-mail: olejnikaa@gmail.com

АНАЛІЗ ГЛИБОКИХ МОДЕЛЕЙ НЕЙРОННИХ МЕРЕЖ НА БАЗІ ОБМЕЖЕНИХ МАШИН БОЛЬЦМАНА

М.Ю. Пришляк, С.О. Субботін, А.О. Олійник

Проведено аналіз моделей глибоких нейронних мереж та методів їх навчання, в результаті якого запропонована класифікація і наведена порівняльна характеристика, а також виділені переваги та недоліки архітектур глибоких нейронних мереж. Проведено експериментальне дослідження з навчання за допомогою обмеженої і глибокої машини Больцмана, а також глибокої мережі довіри для вирішення задачі класифікації на прикладі завдання медичного діагностування раку грудей із застосуванням відкритої бібліотеки DeepLearning4J для побудови нейромоделей. Проведені експерименти дозволили отримати залежності витрат пам'яті ЕОМ програмних реалізацій досліджуваних методів від розмірності вибірок, а також сформулювати рекомендації щодо вибору архітектури глибоких мереж при вирішенні задач класифікації.

Ключові слова: глибока нейронна мережа, машинне навчання, машина Больцмана, класифікація.

ANALYSIS OF DEEP NEURAL NETWORK MODELS BASED ON RESTRICTED BOLTZMANN MACHINES

M. Pryshliak, S. Subbotin, A. Oliinyk

A number of modern approaches exist to perform classification task – Bayes classification, decision trees, artificial neural networks and neuro-fuzzy models. Deep neural networks are the most promising out of them to study, as they can outperform other methods in a lower resulting model complexity, ability to learn from unlabeled data and modeling non-linear distributions within certain limits. This paper aims to perform analysis of deep neural network models based on restricted Boltzmann machines and their learning methods and proposes classification and comparison of their features. It highlights the advantages and disadvantages of deep neural network architectures as well.

In the experiment, restricted and deep Boltzmann machine models were studied, as well as deep belief networks. Their performance was compared on a common classification task with a numeric data arrays – medical diagnosis of breast cancer. DeepLearning4J library was used for building and learning neural networks.

Experiment results made it possible to obtain computer memory usage and neural network performance values of studied learning methods for different deep models. We emphasize recommendations on choosing the correct deep neural network architecture for solving classification problems.

The conducted experiments showed that among the considered network models the most rapidly trained is the restricted Boltzmann machine, the most accurate is the deep belief network. The restricted Boltzmann machine can be effectively used in those cases where the speed of training is important, but not accuracy. Among the examined deep models, a belief network is most effective both in terms of the training speed and of assessing the probability of right decision making.

Keywords: deep neural networks, machine learning, Boltzmann machine, classification.