

УДК 004.89.032.26  
ББК 32.813  
Г63

**Гольдберг Й.**

Г63 Нейросетевые методы в обработке естественного языка / пер. с англ. А. А. Слинкина. – М.: ДМК Пресс, 2019. – 282 с.: ил.

**ISBN 978-5-97060-754-1**

Это классическое руководство посвящено применению нейросетевых моделей к обработке данных естественного языка (Natural Language Processing – NLP). Рассматриваются основы машинного обучения с учителем на лингвистических данных и применение векторных, а не символических представлений слов. Обсуждается абстракция графа вычислений, которая позволяет легко определять и обучать произвольные нейронные сети и лежит в основе современных программных нейросетевых библиотек. Также даются обзорные сведения специализированных нейросетевых архитектур, включая одномерные сверточные сети, рекуррентные нейронные сети, модели условной генерации и модели с механизмом внимания.

Издание предназначено студентам вузов, а также специалистам в области машинного перевода и нейронных сетей.

Предполагается знание теории вероятностей, алгебры и математического анализа, а также базовое владение алгоритмами и структурами данных.

УДК 004.89.032.26  
ББК 32.813

Original English language edition published in the Morgan & Claypool Publishers series. Copyright © 2017 by Morgan & Claypool. Russian-language edition copyright © 2019 by DMK Press. All rights reserved.

Все права защищены. Любая часть этой книги не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами без письменного разрешения владельцев авторских прав.

ISBN 978-1-62705-298-6 (анг.)  
ISBN 978-5-97060-754-1 (рус.)

Copyright © 2017 by Morgan & Claypool  
© Оформление, издание, перевод, ДМК Пресс, 2019

# Содержание

<b>Об авторе .....</b>	<b>12</b>
<b>Предисловие .....</b>	<b>13</b>
<b>Благодарности.....</b>	<b>16</b>
<b>Предисловие от издательства .....</b>	<b>18</b>
<b>Глава 1. Введение .....</b>	<b>19</b>
1.1. Проблемы, стоящие перед обработкой естественного языка.....	19
1.2. Нейронные сети и глубокое обучение.....	20
1.3. Глубокое обучение в NLP .....	21
1.3.1. Истории успеха.....	22
1.4. Состав и организация материала.....	24
1.5. Что не вошло в книгу .....	27
1.6. Замечание о терминологии.....	27
1.7. Математическая нотация .....	28
<b>Часть I. КЛАССИФИКАЦИЯ С УЧИТЕЛЕМ И НЕЙРОННЫЕ СЕТИ ПРЯМОГО РАСПРОСТРАНЕНИЯ .....</b>	<b>29</b>
<b>Глава 2. Основы обучения и линейные модели .....</b>	<b>30</b>
2.1. Обучение с учителем и параметрические функции.....	30
2.2. Обучающий, тестовый и контрольный наборы .....	31
2.3. Линейные модели .....	33
2.3.1. Бинарная классификация.....	33
2.3.2. Логлинейная бинарная классификация .....	37
2.3.3. Многоклассовая классификация.....	38
2.4. Представления .....	39
2.5. Представления в виде унитарного и плотного векторов.....	40
2.6. Логлинейная многоклассовая классификация .....	41
2.7. Обучение как оптимизация.....	41
2.7.1. Функции потерь.....	43
2.7.2. Регуляризация .....	46
2.8. Градиентная оптимизация.....	47
2.8.1. Стохастический градиентный спуск.....	48
2.8.2. Полный пример.....	50
2.8.3. Не только СГС .....	52

<b>Глава 3. От линейных моделей к многослойным перцептронам</b>	<b>53</b>
3.1. Ограничения линейных моделей: проблема XOR	53
3.2. Нелинейные преобразования входа	54
3.3. Ядерные методы	54
3.4. Обучаемые отображающие функции	55
<b>Глава 4. Нейронные сети прямого распространения</b>	<b>56</b>
4.1. Метафора, инспирированная мозгом	56
4.2. Математическая нотация	58
4.3. Репрезентативная способность	60
4.4. Стандартные нелинейности	60
4.5. Функции потерь	62
4.6. Регуляризация и прореживание	62
4.7. Слои вычисления сходства и расстояния	63
4.8. Слои погружения	64
<b>Глава 5. Обучение нейронной сети</b>	<b>65</b>
5.1. Абстракция графа вычислений	65
5.1.1. Прямое вычисление	67
5.1.2. Вычисление на обратном проходе (производные, обратное распространение)	67
5.1.3. Программное обеспечение	68
5.1.4. Рецепт реализации	71
5.1.5. Композиция сети	72
5.2. Практические вопросы	72
5.2.1. Выбор алгоритма оптимизации	72
5.2.2. Инициализация	72
5.2.3. Перезапуск и ансамбли	73
5.2.4. Исчезающие и взрывающиеся градиенты	74
5.2.5. Насыщение и мертвые нейроны	74
5.2.6. Тасование	75
5.2.7. Скорость обучения	75
5.2.8. Мини-пакеты	75
<b>Часть II. РАБОТА С ДАННЫМИ ЕСТЕСТВЕННОГО ЯЗЫКА</b>	<b>76</b>
<b>Глава 6. Признаки для текстовых данных</b>	<b>77</b>
6.1. Типология проблем классификации NLP	77
6.2. Признаки для проблем NLP	79
6.2.1. Непосредственно наблюдаемые свойства	80

6.2.2. Производные лингвистические свойства.....	83
6.2.3. Базовые и комбинационные признаки .....	87
6.2.4. N-граммные признаки .....	87
6.2.5. Дистрибутивные признаки .....	88

## **Глава 7. Примеры признаков в NLP ..... 90**

7.1. Классификация документов: определение языка.....	90
7.2. Классификация документов: тематическая классификация .....	91
7.3. Классификация документов: установление авторства.....	91
7.4. Слово в контексте: частеречная разметка .....	92
7.5. Слово в контексте: распознавание именованных сущностей .....	94
7.6. Слово в контексте, лингвистические признаки: разрешение лексической многозначности предлогов.....	95
7.7. Отношения между словами в контексте: анализ методом разложения на дуги .....	97

## **Глава 8. От текстовых признаков к входным данным ..... 100**

8.1. Кодирование категориальных признаков.....	100
8.1.1. Унитарное кодирование .....	100
8.1.2. Плотные погружения (погружения признаков).....	101
8.1.3. Плотные векторы и унитарные представления.....	101
8.2. Объединение плотных векторов.....	103
8.2.1. Оконные признаки .....	103
8.2.2. Переменное число признаков: непрерывный мешок слов.....	104
8.3. Соотношение между унитарным и плотным векторами.....	105
8.4. Разные разности.....	106
8.4.1. Дистанционные и позиционные признаки .....	106
8.4.2. Дополнение, неизвестные слова и прореживание слов .....	107
8.4.3. Комбинации признаков.....	108
8.4.4. Обобществление векторов .....	109
8.4.5. Размерность.....	110
8.4.6. Словарь погружений .....	110
8.4.7. Выход сети .....	110
8.5. Пример: частеречная разметка.....	111
8.6. Пример: анализ методом разложения на дуги .....	112

## **Глава 9. Языковое моделирование..... 115**

9.1. Задача языкового моделирования.....	115
9.2. Оценивание языковых моделей: перплексивность .....	116
9.3. Традиционные подходы к языковому моделированию.....	117
9.3.1. Для дальнейшего чтения .....	118
9.3.2. Ограничения традиционных языковых моделей .....	118
9.4. Нейросетевые языковые модели .....	119
9.5. Использование языковых моделей для порождения .....	123
9.6. Побочный продукт: представления слов.....	124

<b>Глава 10. Предобученные представления слов .....</b>	<b>125</b>
10.1. Случайная инициализация.....	125
10.2. Специализированное предобучение с учителем.....	125
10.3. Предобучение без учителя .....	126
10.3.1. Использование предобученных погружений.....	127
10.4. Алгоритмы погружения слов.....	128
10.4.1. Дистрибутивная гипотеза и представления слов .....	128
10.4.2. От нейросетевых языковых моделей к распределенным представлениям.....	133
10.4.3. Объединяя миры .....	136
10.4.4. Другие алгоритмы .....	137
10.5. Выбор контекстов.....	138
10.5.1. Подход на основе окон.....	138
10.5.2. Предложения, абзацы или документы .....	139
10.5.3. Синтаксическое окно .....	139
10.5.4. Многоязычные контексты.....	141
10.5.5. Представления на основе литер и подслов .....	141
10.6. Обработка многословных единиц и словоизменения.....	142
10.7. Ограничения дистрибутивных методов .....	143
 <b>Глава 11. Использование погружений слов.....</b>	<b>146</b>
11.1. Получение векторов слов .....	146
11.2. Сходство слов.....	147
11.3. Кластеризация слов.....	147
11.4. Нахождение похожих слов.....	147
11.4.1. Сходство с группой слов .....	148
11.5. Вычеркивание лишних .....	148
11.6. Сходство коротких документов .....	148
11.7. Словесные аналоги .....	149
11.8. Донастройка и проекции.....	150
11.9. Практические вопросы и подводные камни.....	151
 <b>Глава 12. Пример: применение архитектуры прямого распространения для вывода смысла предложения .....</b>	<b>152</b>
12.1. Естественно-языковой вывод и набор данных NLI .....	152
12.2. Сеть для установления сходства текстов .....	154
 <b>Часть III. СПЕЦИАЛИЗИРОВАННЫЕ АРХИТЕКТУРЫ.....</b>	<b>157</b>
 <b>Глава 13. Детекторы <math>n</math>-грамм: сверточные нейронные сети.....</b>	<b>159</b>
13.1. Свертка + пулинг – основы .....	161
13.1.1. Одномерная свертка текста.....	161

13.1.2. Пулинг векторов .....	163
13.1.3. Вариации .....	166
13.2. Альтернатива: хеширование признаков .....	166
13.3. Иерархические свертки .....	167

## Глава 14. Рекуррентные нейронные сети:

<b>последовательности и стеки .....</b>	<b>171</b>
14.1. Абстракция РНС .....	172
14.2. Обучение РНС .....	174
14.3. Типичные примеры использования РНС .....	175
14.3.1. Приемщик .....	175
14.3.2. Кодировщик .....	176
14.3.3. Преобразователь .....	176
14.4. Двухнаправленные РНС (biRNN) .....	177
14.5. Многослойные РНС .....	179
14.6. РНС для представления стеков .....	180
14.7. Замечание о чтении литературы .....	182

## Глава 15. Конкретные архитектуры рекуррентных

<b>нейронных сетей .....</b>	<b>184</b>
15.1. CBOW как РНС .....	184
15.2. Простая РНС .....	184
15.3. Вентильные архитектуры .....	185
15.3.1. LSTM .....	187
15.3.2. GRU .....	188
15.4. Другие варианты .....	189
15.5. Прореживание в РНС .....	190

## Глава 16. Моделирование с помощью рекуррентных

<b>сетей .....</b>	<b>192</b>
16.1. Приемщики .....	192
16.1.1. Классификация по эмоциональной окраске .....	192
16.1.2. Определение грамматической правильности согласования глагола с субъектом .....	194
16.2. РНС как экстракторы признаков .....	196
16.2.1. Частеречная разметка .....	196
16.2.2. Классификация документов с применением РНС-СНС .....	199
16.2.3. Анализ зависимостей методом разложения на дуги .....	199

## Глава 17. Условная генерация .....

17.1. РНС-генераторы .....	202
17.1.1. Обучение генераторов .....	203

17.2. Условная генерация (кодировщик–декодер) .....	203
17.2.1. Модели типа последовательность–в–последовательность .....	205
17.2.2. Приложения .....	207
17.2.3. Другие обуславливающие контексты .....	209
17.3. Установление сходства предложений без учителя .....	210
17.4. Условная генерация с вниманием .....	212
17.4.1. Вычислительная сложность .....	215
17.4.2. Возможность интерпретации .....	215
17.5. Модели на основе внимания в NLP .....	215
17.5.1. Машинный перевод .....	215
17.5.2. Морфологическое словоизменение .....	217
17.5.3. Синтаксический анализ .....	218

## **Часть IV. ДОПОЛНИТЕЛЬНЫЕ ТЕМЫ ..... 219**

### **Глава 18. Моделирование деревьев с помощью рекурсивных нейронных сетей ..... 220**

18.1. Формальное определение .....	221
18.2. Обобщения и вариации .....	223
18.3. Обучение рекурсивных нейронных сетей .....	224
18.4. Простая альтернатива – линейаризованные деревья .....	224
18.5. Перспективы .....	225

### **Глава 19. Предсказание структурного выхода ..... 226**

19.1. Структурное предсказание на основе поиска .....	226
19.1.1. Структурное предсказание с помощью линейных моделей .....	226
19.1.2. Нелинейное структурное предсказание .....	227
19.1.3. Вероятностная целевая функция (CRF) .....	229
19.1.4. Приближенный поиск .....	230
19.1.5. Переранжирование .....	230
19.1.6. Смотрите также .....	231
19.2. Жадное структурное предсказание .....	231
19.3. Условная генерация как предсказание структурного выхода .....	232
19.4. Примеры .....	233
19.4.1. Структурное предсказание на основе поиска: анализ зависимостей первого порядка .....	233
19.4.2. Нейросетевые CRF для распознавания именованных сущностей .....	235
19.4.3. Аппроксимация CRF в задаче NER лучевым поиском .....	238

### **Глава 20. Обучение каскадное, многозадачное и с частичным привлечением учителя ..... 240**

20.1. Каскадирование моделей .....	241
20.2. Многозадачное обучение .....	244

20.2.1. Обучение в многозадачной конфигурации.....	246
20.2.2. Избирательное обобществление.....	246
20.2.3. Предобучение погружений слов как многозадачное обучение.....	247
20.2.4. Многозадачное обучение в условной генерации.....	248
20.2.5. Многозадачное обучение как регуляризация .....	248
20.2.6. Подводные камни .....	248
20.3. Обучение с частичным привлечением учителя.....	249
20.4. Примеры .....	250
20.4.1. Предсказание взгляда и сжатие предложений.....	250
20.4.2. Пометка дуг и синтаксический разбор .....	251
20.4.3. Разрешение лексической многозначности предлогов и предсказание перевода предлогов.....	252
20.4.4. Условная генерация: многоязычный машинный перевод, синтаксический анализ и описание изображений .....	253
20.5. Перспективы.....	254

## **Глава 21. Заключение.....255**

21.1. Что мы узнали?.....	255
21.2. Что ждет впереди?.....	255

## **Список литературы .....257**