

knihovna programátora

- Výklad základních principů hlubokého učení i pokročilých dovedností
- Tvorba systému hlubokého učení pro počítačové vidění, časové řady, text i generování vlastních výtvorů (například obrázků)
- **Způsob fungování moderních AI systémů typu ChatGPT**
- Popis rozdílů při spouštění programů na CPU, GPU a FPU
- Práce s webovým prostředím Collaboration, které umožňuje používat GPU a FPU na serveru



FRANÇOIS CHOLLET

Deep learning

v jazyku Python



2., přepracované a rozšířené vydání

KNIHOVNY KERAS, TENSORFLOW





knihovna programátora

FRANÇOIS CHOLLET

Deep learning v jazyku Python

KNIHOVNY KERAS, TENSORFLOW

2., přepracované a rozšířené vydání

GRADA
Publishing

Upozornění pro čtenáře a uživatele této knihy

Všechna práva vyhrazena. Žádná část této tištěné či elektronické knihy nesmí být reprodukována a šířena v papírové, elektronické či jiné podobě bez předchozího písemného souhlasu nakladatele. Neoprávněné užití této knihy bude **restně stíháno**.

François Chollet

Deep Learning v jazyku Python

Knihovny Keras, TensorFlow

2., přepracované a rozšířené vydání

Přeloženo z anglického originálu knihy Françoise Cholleta Deep Learning with Python, vydaného v roce 2021 nakladatelstvem Manning Publications Co, Spojené státy americké.

Authorized translation of the English edition © 2021 Manning Publications.
This translation is published and sold by permission of Manning Publications,
the owner of all rights to publish and sell the same.
All rights reserved.

Vydala Grada Publishing, a.s.
U Průhonu 22, Praha 7
obchod@grada.cz, www.grada.cz
tel.: +420 234 264 401
jako svou 8701. publikaci

Přeložil: Rudolf Pecinovský
Odpovědný redaktor: Petr Somogyi
Návrh vnitřního layoutu a zlom: Rudolf Pecinovský
Počet stran 528
První vydání, Praha 2023
Vytiskla TISKÁRNA V RÁJI, s.r.o., Pardubice

© Grada Publishing, a.s., 2023
Cover Design © Grada Publishing, a. s., 2023
Cover Photo © Depositphotos/Jirsak

Názvy produktů, firem apod. použité v knize mohou být ochrannými známkami nebo registrovanými ochrannými známkami příslušných vlastníků.

ISBN 978-80-271-7081-4 (pdf)
ISBN 978-80-271-5133-2 (print)

Stručný obsah

Stručný obsah	5
Podrobný obsah	6
Předmluva	14
Poděkování	16
O knize	17
O autorovi	21
O obálce	22
Kapitola 1 Co je hluboké učení	24
Kapitola 2 Než začneme: matematické stavební bloky neuronových sítí	51
Kapitola 3 Úvod do Keras a TensorFlow	93
Kapitola 4 Začínáme s neuronovými sítěmi: Klasifikace a regrese	121
Kapitola 5 Základy strojového učení	148
Kapitola 6 Univerzální pracovní postup strojového učení	181
Kapitola 7 Ponořme se do práce s Keras	201
Kapitola 8 Úvod do hlubokého učení pro počítačové vidění	233
Kapitola 9 Pokročilé hluboké učení pro počítačové vidění	270
Kapitola 10 Hluboké učení pro časové řady	313
Kapitola 11 Hluboké učení pro text	343
Kapitola 12 Generativní hluboké učení	402
Kapitola 13 Osvědčené postupy pro reálný svět	453
Kapitola 14 Závěry	474
Příloha A Terminologický slovník	512
Rejstřík	523

Podrobný obsah

Stručný obsah	5
Podrobný obsah	6
Předmluva	14
Poděkování	16
O knize	17
Komu je kniha určena	17
O kódu18	
Diskusní fórum liveBook.....	19
Použité typografické konvence	19
Odbočka – podšeděný blok.....	20
O autorovi	21
O obálce	22
Kapitola 1 Co je hluboké učení	24
1.1 Umělá inteligence, strojové učení a hluboké učení.....	24
1.1.1 Umělá inteligence (artificial intelligence).....	25
1.1.2 Strojové učení.....	26
1.1.3 Učení se pravidel a reprezentaci z dat	28
1.1.4 Hloubka v hlubokém učení.....	30
1.1.5 Pochopení toho, jak hluboké učení funguje, ve třech krocích	32
1.1.6 Čeho hluboké učení dosud dosáhlo.....	34
1.1.7 Nevěřte krátkodobému hmbuku	35
1.1.8 Příslib AI	36
1.2 Před hlubokým učením: stručná historie strojového učení	37
1.2.1 Pravděpodobnostní modelování.....	38
1.2.2 Rané neuronové sítě	38
1.2.3 Jádrové metody (kernel methods).....	38
1.2.4 Rozhodovací stromy, náhodné lesy a stroje na posílení gradientu.....	40
1.2.5 Zpět k neuronovým sítím.....	41
1.2.6 Co dělá hluboké učení odlišným	42
1.2.7 Krajina moderního strojového učení.....	43
1.3 Proč hluboké učení? Proč teď?	45
1.3.1 Hardware	45
1.3.2 Data.....	46
1.3.3 Algoritmy	47
1.3.4 Nová vlna investic	48
1.3.5 Demokratizace hlubokého učení	49
1.3.6 Vydrží to?	49
Kapitola 2 Než začneme: matematické stavební bloky neuronových sítí.....	51
2.1 První pohled na neuronovou síť	52
2.2 Reprezentace dat pro neuronové sítě.....	56
2.2.1 Skaláry (tenzory nultého řádu, 0D tenzory).....	56

2.2.2	Vektory (tenzory 1. řádu, 1D tenzory)	57
2.2.3	Matice (tenzory 2. řádu, 2D tenzory)	57
2.2.4	Tenzory tří a vícedimenzionální	57
2.2.5	Klíčové atributy	58
2.2.6	Manipulace s tenzory v <i>NumPy</i>	59
2.2.7	Pojem dávek dat	60
2.2.8	Příklady datových tenzorů v reálném světě	60
2.2.9	Vektorová data	61
2.2.10	Časové řady nebo sekvenční data	61
2.2.11	Obrazová data	62
2.2.12	Video data	62
2.3	Nástroje neuronových sítí: tenzorové operace	63
2.3.1	Elementové operace (element-wise operations)	64
2.3.2	Vysílání (broadcasting)	65
2.3.3	Tenzorový součin	66
2.3.4	Změna tvaru (přetváření) tenzoru (tensor reshaping)	69
2.3.5	Geometrická interpretace tenzorových operací	69
2.3.6	Geometrická interpretace hlubokého učení	73
2.4	Motor neuronových sítí: optimalizace založená na gradientu	73
2.4.1	Co je derivace?	75
2.4.2	Derivace tenzorové operace: gradient	76
2.4.3	Stochastický gradientní sestup	77
2.4.4	Zřetězení derivací: algoritmus zpětného šíření	80
	Pravidlo řetězu	81
	Automatická derivace s výpočetními grafy	81
	GradientTape v TensorFlow	86
2.5	Ohlédnutí za naším prvním příkladem	86
2.5.1	Reimplementace prvního příkladu od nuly v TensorFlow	88
	Jednoduchá třída NaiveDense	88
	Jednoduchá sekvenční třída	89
	Generátor dávek	89
2.5.2	Běh jednoho tréninkového kroku	90
2.5.3	Úplný trénovací cyklus	91
2.5.4	Vyhodnocení modelu	91
2.6	Shrnutí kapitoly	92
Kapitola 3 Úvod do Keras a TensorFlow		93
3.1	Co je TensorFlow	93
3.2	Co je Keras	94
3.3	Keras a TensorFlow: stručná historie	96
3.4	Nastavení pracovního prostoru pro hluboké učení	96
3.4.1	Notebooky Jupyter: preferovaný způsob provádění experimentů s hlubokým učením	97
3.4.2	Používání laboratoře Colaboratory	98
	První kroky s Colaboratory	98
	Instalace balíčků pomocí PIP	100
	Použití běhového prostředí GPU	100
3.5	První kroky s TensorFlow	101
3.5.1	Konstantní tenzory a proměnné	101
3.5.2	Operace s tenzory: provádění matematických operací v TensorFlow	104
3.5.3	Druhý pohled na API GradientTape	104
3.5.4	Souhrnný příklad: lineární klasifikátor v čistém TensorFlow	105
3.6	Anatomie neuronové sítě: pochopení základních rozhraní Keras API	110
3.6.1	Vrstvy: základní kameny hlubokého učení	110
	Layer – základní třída vrstvy v <i>Keras</i>	111
	Automatické odvozování tvaru: vytváření vrstev za běhu	112
3.6.2	Od vrstev k modelům	113

3.6.3	Krok „překlad“: konfigurace procesu učení.....	115
3.6.4	Výběr ztrátové funkce.....	117
3.6.5	Porozumění metodě <code>fit()</code>	117
3.6.6	Monitorování ztrát a metrik na validačních datech.....	118
3.6.7	Odvození: použití modelu po tréninku.....	119
3.7	Shrnutí kapitoly.....	120
Kapitola 4	Začínáme s neuronovými sítěmi: Klasifikace a regrese	121
	Glosář klasifikace a regrese.....	121
4.1	Klasifikace filmových recenzí: příklad binární klasifikace	123
4.1.1	Databáze IMDB.....	123
4.1.2	Příprava dat.....	124
4.1.3	Sestavení modelu.....	125
	Co jsou aktivační funkce a proč jsou potřebné.....	128
4.1.4	Ověření vašeho přístupu.....	128
4.1.5	Použití vycvičeného modelu k vytváření předpovědí na nových datech.....	131
4.1.6	Další experimenty.....	132
4.1.7	Shrnutí.....	132
4.2	Klasifikace zpravodajství: příklad klasifikace do více tříd	133
4.2.1	Soubor dat <i>Reuters</i>	133
4.3.2	Příprava dat.....	134
4.2.3	Sestavení modelu.....	135
4.2.4	Ověření vašeho přístupu.....	136
4.2.5	Generování predikcí pro nová data.....	138
4.2.6	Jiný způsob zacházení se štítky a ztrátou.....	138
4.2.7	Důležitost dostatečně velkých mezivrstev.....	139
4.2.8	Další experimenty.....	139
4.2.9	Shrnutí.....	139
4.3	Predikce cen nemovitostí: příklad regrese	140
4.3.1	Soubor údajů o cenách bydlení v Bostonu.....	140
4.3.2	Příprava dat.....	141
4.3.3	Sestavení modelu.....	141
4.3.4	Ověření vašeho přístupu použitím k-násobné validace.....	142
4.3.5	Generování předpovědí na nových datech.....	146
4.3.5	Shrnutí.....	146
4.4	Shrnutí kapitoly.....	147
Kapitola 5	Základy strojového učení	148
5.1	Zobecnění: cíl strojového učení	148
5.1.1	Přeučení a podučení	149
	Zašuměná trénovací data.....	149
	Nejednoznačné rysy.....	151
	Vzácné rysy a falešné korelace.....	151
5.1.2	Povaha zobecnění v hlubokém učení	154
	Hypotéza o mnohotvárnosti.....	155
	Interpolace jako zdroj zobecnění.....	156
	Proč hluboké učení funguje.....	157
	Trénovací data jsou nejdůležitější.....	158
5.2	Vyhodnocování modelů strojového učení	160
5.2.1	Trénovací, validační a testovací sady	160
	Jednoduchá validace na odebraných datech (simple hold-out validation).....	161
	K-násobné křížová validace.....	162
	Opakovaná k-násobná validace s promícháním.....	163
5.2.2	Překonání základní referenční úrovně	163
5.2.2	Věci, které je třeba mít na paměti	164
5.3	Zlepšení odpovědi modelu	165
5.3.1	Ladění klíčových parametrů gradientního sestupu	165
5.3.2	Využití lepších předloh architektury	166

5.3.3 Zvyšování kapacity modelu	167
5.4 Zlepšení zobecnění	169
5.4.1 Kurátorství datových sad	169
5.4.2 Inženýrství rysů	170
5.4.3 Včasné zastavení	171
5.4.4 Regularizace modelu	172
Zmenšení velikosti sítě	172
Přidání regularizace vah	175
Přidání dropout	177
5.5 Shrnutí kapitoly	179
Kapitola 6 Univerzální pracovní postup strojového učení	181
Poznámka k etice	182
6.1 Definice úkolu	183
6.1.1 Rámcový popis problému	183
6.1.2 Shromáždění souboru dat	185
Investice do infrastruktury pro anotaci dat	186
Pozor na nereprezentativní údaje	186
Problém zkeslení výběru vzorku	187
6.1.3 Porozumění datům	188
6.1.4 Zvolte si měřítko úspěchu	189
6.2 Vyvinutí modelu	189
6.2.1 Příprava dat	189
Vektorizace	190
Normalizace hodnot	190
Zpracování chybějících hodnot	190
6.2.2 Výběr hodnotícího protokolu	191
6.2.3 Překonání základní referenční úrovně	191
Výběr správné ztrátové funkce	192
6.2.4 Rozšíření: vývoj modelu, který se přizpůsobí	193
6.2.5 Regularizace a vyladění modelu	193
6.3 Nasazení modelu	194
6.3.1 Vysvětlete svou práci zúčastněným stranám a stanovte očekávání	194
6.3.2 Odeslání odvozovacího modelu	195
Nasazení modelu jako REST API	195
Nasazení modelu v zařízení	196
Nasazení modelu v prohlížeči	197
Optimalizace odvozovacího (inferenčního) modelu	198
6.3.3 Sledování modelu „v divočině“	198
6.3.4 Udržujte svůj model	199
6.4 Shrnutí kapitoly	199
Kapitola 7 Ponořme se do práce s Keras	201
7.1 Spektrum pracovních postupů	202
7.2 Různé způsoby vytváření modelů Keras	202
7.2.1 Sekvenční model	203
7.2.2 Funkcionální API	205
Jednoduchý příklad	205
Modely s více vstupy a výstupy	207
Trénování modelu s více vstupy a výstupy	208
Síla funkcionálního API: přístup k propojení vrstev	209
7.2.3 Definice potomků třídy Model	211
Přepsání předchozího příkladu jako modelu s podtřídou	211
Pozor: co modely s podtřídou nepodporují	213
7.2.4 Míchání a kombinování různých součástí	213
7.2.5 Pamatujte: pro danou práci použijte správný nástroj	214
7.3 Použití vestavěných trénovacích a vyhodnocovacích cyklů	215
7.3.1 Psaní vlastních metrik	216

7.3.2	Použití zpětných volání.....	217
	Zpětná volání <code>EarlyStopping</code> a <code>ModelCheckpoint</code>	218
7.3.3	Psaní vlastních zpětných volání	219
7.3.4	Monitorování a vizualizace pomocí TensorBoard.....	221
7.4	Psaní vlastních školicích a hodnoticích cyklů.....	223
	Další druhy učení.....	224
7.4.1	Skolení versus odvozování	225
7.4.2	Použití metrik na nízké úrovni.....	226
7.4.3	Kompletní cyklus školení a hodnocení	226
7.4.4	Zrychlete pomocí funkce <code>tf.function</code>	228
7.4.5	Využití funkce <code>fit()</code> s vlastním trénovacím cyklem	229
7.5	Shrnutí kapitoly.....	231
Kapitola 8	Úvod do hlubokého učení pro počítačové vidění	233
8.1	Úvod do konvolučních neuronových sítí – CNN.....	234
8.1.1	Konvoluční operace	236
	Porozumění hraničním efektům a vycpávkám	239
	Porozumění konvolučním krokům	240
8.1.2	Operace sdružování dle maxima (max pooling).....	241
8.2	Trénování CNN od nuly na malé sadě dat	243
8.2.1	Význam hlubokého učení pro úlohy s malými daty	243
8.2.2	Stážení dat.....	244
	Stážení datové sady Kaggle v nástroji Google Colaboratory	245
8.2.3	Sestavení modelu	247
8.2.4	Předzpracování dat	249
	Porozumění objektům <code>Dataset</code> frameworku TensorFlow	249
8.2.5	Použití rozšíření dat (data augmentation).....	253
8.3	Použití předtrénované CNN	256
8.3.1	Extrakce rysů	257
	Rychlá extrakce rysů bez rozšíření dat	261
	Extrakce rysů s rozšířením dat	263
8.3.2	Jemné doladění předtrénovaného modelu	265
8.4	Shrnutí kapitoly.....	268
Kapitola 9	Pokročilé hluboké učení pro počítačové vidění	270
9.1	Tři základní úlohy počítačového vidění.....	270
9.2	Příklad segmentace obrazu	272
9.3	Moderní vzory architektury CNN	280
9.3.1	Modularita, hierarchie a opakované použití	281
	O významu ablačních studií ve výzkumu hlubokého učení	283
9.3.2	Zbytková připojení (residual connections).....	283
9.3.3	Dávková normalizace.....	287
9.3.4	Hloubkově oddělitelné konvoluce	289
	Koevoluce hardwaru, softwaru a algoritmů.....	290
9.3.5	Složení: mini model podobný Xception	292
9.4	Interpretace toho, co se CNN naučí	294
9.4.1	Vizualizace mezilehlých aktivací.....	294
9.4.2	Vizualizace filtrů CNN	300
	Rozdíl mezi <code>model.predict(x)</code> a <code>model(x)</code>	302
9.4.3	Zobrazení teplotních map aktivací třídy.....	306
9.5	Shrnutí kapitoly.....	312
Kapitola 10	Hluboké učení pro časové řady	313
10.1	Různé druhy úloh časových řad	313
10.2	Příklad předpovědi teploty.....	314
	Vždy hledejte v datech periodicitu	317
10.2.1	Příprava dat.....	317
	Porozumění <code>timeseries_dataset_from_array()</code>	318

10.2.2	Základní referenční úroveň bez strojového učení.....	320
10.2.3	Vyzkoušejme základní model strojového učení	321
10.2.4	Vyzkoušejme 1D konvoluční model.....	323
10.2.5	První rekurentní základní referenční úroveň	325
10.3	Porozumění rekurentním neuronovým sítím	326
10.3.1	Rekurentní vrstvy v <i>Keras</i>	329
10.4	Pokročilé využití rekurentních neuronových sítí	333
10.4.1	Použití rekurentního dropoutu v boji proti přeučení.....	334
	Výkonnost RNN za běhu	335
10.4.2	Stohování rekurentních vrstev	336
10.4.3	Použití obousměrných RNN	337
10.4.4	Pokračujme ještě dále.....	340
	Trhy a strojové učení.....	341
10.5	Shrnutí kapitoly.....	341
Kapitola 11	Hluboké učení pro text.....	343
11.1	Zpracování přirozeného jazyka: z ptací perspektivy.....	343
11.2	Příprava textových dat	345
11.2.1	Standardizace textu	346
11.2.2	Dělení textu (tokenizace).....	347
	Porozumění N-gramům a sadám slov.....	348
11.2.3	Indexování slovní zásoby	348
11.2.4	Použití vrstvy <i>TextVectorization</i>	350
	Použití vrstvy <i>TextVectorization</i> v datovodu <i>tf.data</i> nebo jako součást modelu	352
11.3	Dva přístupy k reprezentaci skupin slov: sady a posloupnosti	354
11.3.1	Příprava dat recenzí filmů na IMDB.....	354
11.3.2	Zpracování slov jako sady: přístup sada slov (Bag-of-Words approach)	356
	Jednotlivá slova s binárním kódováním (1-gramy, unigramy).....	357
	Bigramy (2-gramy) s binárním kódováním	359
	Bigramy s kódováním TF-IDF	360
	Porozumění normalizaci TF-IDF	361
	Export modelu, který zpracovává nezpracované řetězce.....	362
11.3.3	Zpracování slov jako sekvence: přístup sekvencního modelu.....	363
	První praktický příklad	363
	Porozumění slovním vnořením.....	365
	Učení vkládání slov pomocí vrstvy <i>Embedding</i>	367
	Porozumění výplním a maskování	368
	Použití předem natrénovaných slovních vnoření	370
11.4	Architektura Transformátor	372
11.4.1	Porozumění sebezpozornosti.....	373
	Zobecněná sebezpozornost: model dotaz-klíč-hodnota	376
11.4.2	Pozornost více hlav (multi-head attention)	377
11.4.3	Transformátor	378
	Ukládání vlastních vrstev.....	380
	Použití pozičního kódování k opětovnému vložení informací o pořadí	383
	Všechno dohromady: transformátor pro klasifikaci textu	384
11.4.4	Kdy použít sekvencní modely místo modelů typu „sada slov“?	385
11.5	Nad rámec klasifikace textu: učení sekvence na sekvenci	386
11.5.1	Příklad strojového překladu	387
11.5.2	Učení sekvence na sekvenci pomocí RNN.....	390
11.5.3	Učení sekvence na sekvenci pomocí Transformátoru.....	395
	Dekodér Transformátoru	396
	Vše dohromady: Transformátor pro strojový překlad	398
11.6	Shrnutí kapitoly.....	401
Kapitola 12	Generativní hluboké učení	402
12.1	Generování textu	404
12.1.1	Stručná historie generativního hlubokého učení pro generování sekvencí.....	404

12.1.2	Jak generujete sekvenční data?	405
12.1.3	Význam strategie výběru vzorků	406
12.1.4	Implementace generování textu pomocí Keras	407
	Příprava dat	408
	Model založený na transformátoru sekvence-sekvenci	409
12.1.5	Zpětné volání pro generování textu se vzorkováním s proměnnou teplotou	411
12.1.6	Závěrečné shrnutí	415
12.2	DeepDream	415
12.2.1	Implementace DeepDream v Keras	416
12.2.2	Závěrečné shrnutí	422
12.3	Přenos neuronového stylu	423
12.3.1	Ztráta obsahu	424
12.3.2	Ztráta stylu	424
12.3.3	Přenos neuronového stylu v Keras	425
12.3.4	Závěrečné shrnutí	431
12.4	Generování obrazů pomocí variačních autoenkodérů	431
12.4.1	Vzorkování z latentních prostorů obrazů	431
12.4.2	Koncepční vektory pro úpravy obrázků	432
12.4.3	Variační autoenkodéry	433
12.4.4	Implementace VAE pomocí Keras	436
12.4.5	Závěrečné shrnutí	441
12.5	Úvod do generativních soupeřících sítí	442
12.5.1	Schematická implementace GAN	443
12.5.2	Sada triků	444
12.5.3	Získání souboru dat CelebA	445
12.5.4	Diskriminátor	446
12.5.5	Generátor	447
12.5.6	Soupeřící síť	448
12.5.7	Závěrečné shrnutí	451
12.6	Shrnutí kapitoly	452
Kapitola 13	Osvědčené postupy pro reálný svět	453
13.1	Využití vašich modelů na maximum	454
13.1.1	Optimalizace hyperparametrů	454
	Používání KerasTuneru	455
	Maximalizace a minimalizace cíle optimalizace	457
	Umění vytvořit správný vyhledávací prostor	460
	Budoucnost ladění hyperparametrů: automatizované strojové učení	460
13.1.2	Kombinování modelů	461
13.2	Rozšiřování školení modelů	463
13.2.1	Zrychlení tréninku na GPU se smíšenou přesností	464
	Porozumění přesnosti reálných čísel	464
	Poznámka ke kódování reálných čísel	465
	Pozor na výchozí hodnoty dtype	466
	Trénink se smíšenou přesností v praxi	467
13.2.2	Školení s více grafickými procesory	467
	Získání dvou nebo více GPU	468
	Synchronní školení s jedním hostitelem a více zařízeními	468
	Tipy pro výkon tf.data	469
13.2.3	Trénování TPU	470
	Použití TPU přes Google Colab	470
	Pozor na úzká místa I/O	472
	Využití slučování kroků ke zlepšení využití TPU	472
13.3	Shrnutí kapitoly	472
Kapitola 14	Závěry	474
14.1	Přehled klíčových pojmů	474
14.1.1	Různé přístupy k AI	475

14.1.2	Čím je hluboké učení v oblasti strojového učení výjimečné.....	475
14.1.3	Jak uvažovat o hlubokém učení.....	476
14.1.4	Klíčové podpůrné technologie.....	477
14.1.5	Univerzální pracovní postup strojového učení.....	478
14.1.6	Klíčové síťové architektury.....	479
	Hustě propojené sítě.....	480
	CNN (konvoluční neuronové sítě).....	481
	RNN.....	482
	Transformátory.....	482
14.1.7	Prostor možností.....	483
14.2	Omezení hlubokého učení.....	485
14.2.1	Riziko antropomorfizace modelů strojového učení.....	486
14.2.2	Automaty vs. inteligentní agenti.....	488
14.2.3	Lokální generalizace vs. extrémní generalizace.....	489
14.2.4	Účel inteligence.....	491
14.2.5	Šplhání po spektru zobecnění.....	492
14.3	Nastavení směru k větší obecnosti umělé inteligence.....	493
14.3.1	O důležitosti stanovení správného cíle: pravidlo zkratky.....	493
14.3.2	Nový cíl.....	495
14.4	Zavádění inteligence: chybějící ingredience.....	497
14.4.1	Inteligence jako citlivost na abstraktní analogie.....	497
14.4.2	Dva póly abstrakce.....	498
	Analogie zaměřená na hodnoty (value-centric analogy).....	499
	Analogie zaměřená na program (program-centric analogy).....	500
	Poznávání jako kombinace obou druhů abstrakce.....	501
14.4.3	Chybějící polovina obrázku.....	501
14.5	Budoucnost hlubokého učení.....	502
14.5.1	Modely jako programy.....	503
14.5.2	Spojení hlubokého učení a syntézy programů.....	504
	Integrace modulů hlubokého učení a algoritmických modulů do hybridních systémů.....	505
	Využití hlubokého učení k vyhledávání programů.....	506
14.5.3	Celoživotní učení a modulární opakované použití podprogramů.....	507
14.5.4	Dlouhodobá vize.....	508
14.6	Zůstaňte v obraze v rychle se vyvíjející oblasti.....	509
14.6.1	Cvičení na reálných úlohách pomocí Kaggle.....	510
14.6.2	Přečtete si o nejnovějším vývoji na arXiv.....	510
14.6.3	Prozkoumejte ekosystém Keras.....	511
14.7	Závěrečné slovo.....	511
Příloha A	Terminologický slovník.....	512
A.1	Používané zkratky.....	512
	GRU.....	513
A.2	Řazený dle anglických termínů.....	513
A.3	Řazený dle českých termínů.....	518
Rejstřík	523

Předmluva

Držíte-li tuto knihu v ruce, jste si asi vědomi mimořádného pokroku, jehož hluboké učení pro oblast umělé inteligence v nedávné době dosáhlo. Za pouhých pět let jsme se od téměř nepoužitelného počítačového vidění a zpracování přirozeného jazyka dostali až k vysoce výkonným systémům nasazeným ve velkém měřítku v produktech, které používáme každý den.

Důsledky tohoto náhlého pokroku zasahují téměř do všech odvětví. Hluboké učení již uplatňujeme v rozsáhlé škále důležitých problémů v tak odlišných oblastech, jako je lékařské zobrazování, zemědělství, autonomní řízení, vzdělávání, prevence katastrof a výroba.

Přesto se domnívám, že hluboké učení je stále v počátcích. Zatím využilo jen malou část svého potenciálu. Postupem času se dostane do všech oblastí, kde může pomoci – tato transformace bude trvat několik desetiletí.

Abychom mohli začít používat technologii hlubokého učení na všechny problémy, které by mohla vyřešit, musíme ji zpřístupnit co největšímu počtu lidí, včetně těch, kteří nejsou odborníky: tedy i lidem, kteří nejsou výzkumníky nebo postgraduálními studenty. Aby hluboké učení dosáhlo plného potenciálu, musíme ho radikálně demokratizovat.

Domnívám, že nyní jsme na vrcholu historického přechodu, kdy hluboké učení opouští vědecké laboratoře a výzkumná či vývojová oddělení velkých technologických společností a stává se všudypřítomnou součástí sady nástrojů každého vývojáře, ne nepodobně trajektorii vývoje webových stránek na konci 90. let. Téměř každý si nyní může pro svou firmu nebo komunitu vytvořit webové stránky nebo webovou aplikaci, k jejichž vytvoření by v roce 1998 potřeboval malý tým inženýrů specialistů. V nepříliš vzdálené budoucnosti bude moci kdokoli s nápadem a základními dovednostmi v oblasti kódování vytvářet inteligentní aplikace, které se budou učit z dat.

Když jsem v březnu 2015 vydal první verzi frameworku *Keras* pro hluboké učení, neměl jsem na mysli demokratizaci umělé inteligence. Již několik let jsem se zabýval výzkumem v oblasti strojového učení a vytvořil jsem *Keras*, aby mi pomohl při mých vlastních experimentech.

Od roku 2015 však do oblasti hlubokého učení vstoupily statisíce nováčků; mnozí z nich si *Keras* vybrali jako svůj nástroj. Když jsem sledoval, jak desítky chytrých lidí používají *Keras* nečekaným a efektivním způsobem, začal jsem se velmi zajímat o dostupnost a demokratizaci umělé inteligence. Uvědomil jsem si, že čím více se tyto technologie šíří, tím jsou užitečnější a cennější. Přístupnost se rychle stala cílem vývoje

Keras a během několika krátkých let dosáhla komunita vývojářů v tomto směru fantastických úspěchů. Dali jsme hluboké učení do rukou statisíců lidí, kteří ho následně používají k řešení problémů, jež byly donedávna považovány za neřešitelné.

Kniha, kterou držíte v ruce, je dalším krokem na cestě k zpřístupnění hlubokého učení co největšímu počtu lidí. *Keras* vždy potřeboval doprovodný kurz, který by se současně zabýval základy hlubokého učení, osvědčenými postupy a způsoby použití frameworku *Keras*.

V letech 2016 a 2017 jsem se snažil takový kurz vytvořit, což vedlo k první verzi této knihy, která vyšla v prosinci 2017. Rychle se stala bestsellerem v oblasti strojového učení, prodalo se jí přes 50 000 výtisků a byla přeložena do 12 jazyků.

Oblast hlubokého učení však rychle postupuje vpřed. Od prvního vydání došlo k mnoha důležitým změnám – objevilo se *TensorFlow 2*, vzrostla popularita architektury *Transformátor* a mnoho dalšího. A tak jsem se koncem roku 2019 rozhodl svou knihu aktualizovat.

Původně jsem si docela naivně myslel, že bude obsahovat zhruba 50 % nového obsahu, nakonec bude zhruba stejně dlouhá jako první vydání. V praxi se po dvou letech práce ukázalo, že je více než o třetinu delší a obsahuje asi 75 % nového obsahu. Spíše než o aktualizaci tedy jde o zcela novou knihu.

Napsal jsem ji s důrazem na to, aby koncepty hlubokého učení a jejich implementace byly co nejpřístupnější. Přitom jsem nemusel nic hloupě vysvětlovat, protože pevně věřím, že v hlubokém učení nejsou žádné obtížné myšlenky. Doufám, že pro vás bude tato kniha cenná a že vám umožní začít vytvářet inteligentní aplikace a řešit problémy, na kterých vám záleží.

Poznámka překladatele a redakce k názvu knihy: V předkládané publikaci se snažíme maximálně dodržovat českou terminologii, ale u názvu knihy jsme dali přednost anglickému termínu *deep learning*. Domníváme se, že je ve větším povědomí mezi čtenáři než jeho český ekvivalent *hluboké učení*.

Poděkování

Nejprve bych rád poděkoval komunitě *Keras* za to, že umožnila vznik této knihy. Za posledních šest let se *Keras* rozrostl na stovky přispěvatelů do open source a více než milion uživatelů. Díky vašim příspěvkům a zpětné vazbě se *Keras* stal tím, čím je dnes.

Z osobních důvodů bych rád poděkoval své ženě za její nekonečnou podporu při vývoji frameworku *Keras* a psaní této knihy.

Rád bych také poděkoval společnosti *Google* za podporu projektu *Keras*. Bylo fantastické, že *Keras* byl přijat jako vysokoúrovňové API frameworku *TensorFlow*. Hladká integrace mezi frameworky *Keras* a *TensorFlow* je velmi přínosná jak pro uživatele *TensorFlow*, tak pro uživatele *Keras*, a zpřístupňuje hluboké učení většině.

Chtěl bych rovněž poděkovat lidem z nakladatelství *Manning*, kteří se zasloužili o vznik této knihy: nakladateli Marjanu Baceovi a všem členům redakčního a produkčního týmu včetně Michaela Stephense, Jennifer Stoutové, Aleksandara Dragosavljeviče a mnoha dalších, kteří pracovali v zákulisí.

Velký dík patří technickým recenzentům: Billy O'Callaghan, Christian Weisstanner, Conrad Taylor, Daniela Zapata Riesco, David Jacobs, Edmon Begoli, Edmund Ronald, Ph.D., Hao Liu, Jared Duncan, Kee Nam, Ken Fricklas, Kjell Jansson, Milan Šarenac, Nguyen Cao, Nikos Kanakaris, Oliver Korten, Raushan Jha, Sayak Paul, Sergio Govoni, Shashank Polasa, Todd Cook a Viton Vitanis – a všem dalším, kteří nám poslali zpětnou vazbu k návrhu knihy.

Po technické stránce patří zvláštní poděkování Frances Buontempo, technické redaktorce knihy, a Karstenu Strøbækovi, který se postaral o korektury.