

Wydanie 2

Generatywne głębokie uczenie

*Uczenie maszyn, jak malować,
pisać, komponować i grać*

David Foster
Słowo wstępne Karl Friston

przekład: Mariusz Rogulski, Witold Sikorski

Spis treści

Przedmowa	xv
Wstęp	xvii

Część I Wprowadzenie do generatywnego głębokiego uczenia

1. Modelowanie generatywne.....	3
Czym jest modelowanie generatywne?	4
Modelowanie generatywne vs. dyskryminacyjne	5
Powstanie modelowania generatywnego	6
Modelowanie generatywne i sztuczna inteligencja	8
Nasz pierwszy model generatywny	9
Witaj, świecie!	9
Ramy modelowania generatywnego	10
Uczenie reprezentacji	12
Podstawowa teoria prawdopodobieństwa	15
Taksonomia modelu generatywnego	18
Baza kodu dla generatywnego głębokiego uczenia	20
Klonowanie repozytorium	20
Użycie Dockera	20
Uruchamianie na GPU	20
Podsumowanie	21
2. Uczenie głębokie	23
Dane dla uczenia głębokiego	24
Głębokie sieci neuronowe	25
Czym jest sieć neuronowa?	25
Uczenie się wysokopoziomowych cech	26
TensorFlow i Keras	27
Perceptron wielowarstwowy (MLP)	28
Przygotowanie danych	28
Budowanie modelu	30
Kompilowanie modelu	35
Szkolenie modelu	36
Ocena modelu	38
Splotowa sieć neuronowa (CNN)	40

Warstwy splotowe	40
Wsadowa normalizacja	45
Odrzucanie	48
Budowanie CNN	50
Szkolenie oraz ocena CNN	52
Podsumowanie	53

Część II Metody

3. Autokodery wariacyjne	57
Wprowadzenie	58
Autokodery	59
Zbiór danych Fashion-MNIST	60
Architektura autokodera	61
Koder	62
Dekoder	63
Połączenie kodera z dekodere	65
Rekonstruowanie obrazów	67
Wizualizowanie przestrzeni ukrytej	68
Generowanie nowych obrazów	69
Autokodery wariacyjne	71
Koder	72
Funkcja straty	78
Szkolenie autokodera wariacyjnego	79
Analiza autokodera wariacyjnego	81
Eksploracja przestrzeni ukrytej	82
Zbiór danych CelebA	83
Szkolenie autokodera wariacyjnego	84
Analiza autokodera wariacyjnego	86
Generowanie nowych twarzy	87
Arytmetyka przestrzeni ukrytej	88
Morfing pomiędzy twarzami	89
Podsumowanie	90
4. Generatywne sieci antagonistyczne	91
Wprowadzenie	92
Głęboka splotowa sieć GAN (DCGAN)	94
Zbiór danych Bricks	94
Dyskryminator	95
Generator	97
Szkolenie DCGAN	101
Analiza DCGAN	106
Szkolenie GAN: rady i wskazówki	107

GAN Wassersteina z karą gradientową (WGAN-GP)	110
Strata Wassersteina	111
Ograniczenie Lipschitza	112
Wymuszanie ograniczenia Lipschitza	113
Kara gradientowa w stracie	114
Szkolenie sieci WGAN-GP	115
Analiza sieci WGAN-GP	117
Warunkowa GAN (CGAN)	119
Architektura sieci CGAN	119
Szkolenie sieci CGAN	121
Analiza sieci CGAN	123
Podsumowanie	124
5. Modele autoregresyjne	125
Wprowadzenie	126
Sieć z długą pamięcią krótkotrwałą (LSTM)	128
Zbiór danych Recipes	128
Praca z danymi tekstowymi	130
Tokenizacja	131
Tworzenie zbioru szkoleniowego	134
Architektura LSTM	134
Warstwa zagnieżdżeń	135
Warstwa LSTM	136
Komórka LSTM	138
Szkolenie LSTM	140
Analiza LSTM	142
Rozszerzenia rekurencyjnej sieci neuronowej (RNN)	145
Stosowe sieci rekurencyjne	145
Bramkowane jednostki rekurencyjne	147
Komórki dwukierunkowe	148
PixelCNN	149
Maskowane warstwy spłotowe	149
Bloki resztkowe	151
Szkolenie PixelCNN	153
Analiza PixelCNN	154
Mieszana dystrybucja	157
Podsumowanie	159
6. Modele przepływów normalizujących	161
Wprowadzenie	162
Przepływy normalizujące	164
Zamiana zmiennych	164
Wyznacznik Jacobiego	166

Równanie zamiany zmiennych	167
RealNVP	168
Zbiór danych Two Moons	168
Warstwy sprzęgające	169
Szkolenie modelu RealNVP	174
Analiza modelu RealNVP	177
Inne modele przepływu normalizującego	179
GLOW	179
FFJORD	180
Podsumowanie	181
7. Modele oparte o energię	183
Wprowadzenie	184
Modele oparte o energię	185
Zbiór danych MNIST	186
Funkcja energii	187
Próbkowanie z użyciem dynamiki Langevina	189
Szkolenie z dywergencją przeciwną	191
Analiza modeli opartych o energię	195
Inne modele oparte o energię	196
Podsumowanie	197
8. Modele dyfuzji	199
Wprowadzenie	200
Odszumiające modele dyfuzji (DDM)	202
Zbiór danych Flowers	202
Proces dyfuzji w przód	204
Sztuczka reparametryzacyjna	205
Rozkłady dyfuzji	206
Proces dyfuzji odwrotnej	208
Odszumiający model U-Net	211
Szkolenie modelu dyfuzji	218
Próbkowanie z odszumiającego modelu dyfuzji	219
Analiza modelu dyfuzji	222
Podsumowanie	225
 Część III Zastosowania	
9. Transformery	229
Wprowadzenie	230
GPT	230
Zbiór danych z Wine Reviews	231
Uwaga	232

Zapytania, klucze i wartości	233
Uwaga wielogłowicowa	235
Nieformalne maskowanie	236
Blok transformerów	238
Kodowanie pozycyjne	241
Szkolenie GPT	243
Analiza GPT	245
Inny transformer	247
T5	248
GPT-3 i GPT-4	251
ChatGPT	252
Podsumowanie	256
10. Zaawansowane GAN	257
Wprowadzenie	258
ProGAN	258
Szkolenie progresywne	259
Wyniki	265
StyleGAN	265
Sieć mapująca	267
Sieć syntetyzująca	267
Wyniki z StyleGAN	268
StyleGAN2	268
Modulacja i demodulacja wagi	270
Regularyzacja długości ścieżki	271
Brak progresywnego wzrostu	272
Wyniki z StyleGAN2	273
Inne ważne odmiany GAN	274
GAN (SAGAN) skupiony na sobie	274
BigGAN	276
VQ-GAN	277
ViT VQ-GAN	280
Podsumowanie	282
11. Generowanie muzyki	283
Wprowadzenie	284
Transformery do generowania muzyki	285
Zbiór danych suita na wiolonczelę Bacha	286
Parsowanie plików MIDI	286
Tokenizacja	288
Tworzenie zbioru szkoleniowego	289
Sinusoidalne kodowanie pozycji	290
Wiele wejść i wyjść	292

Analiza transformera generującego muzykę	294
Tokenizacja muzyki polifonicznej	298
MuseGAN	302
Zbiór danych z chorałem Bacha	302
Generator MuseGAN	305
Krytyk MuseGAN	311
Analiza MuseGAN	312
Podsumowanie	314
12. Modele światowe	317
Wprowadzenie	318
Uczenie przez wzmacnianie	318
Środowisko CarRacing	320
Ogólne spojrzenie na model światowy	322
Architektura	322
Szkolenie	325
Zbieranie losowo zbieranych danych	325
Szkolenie VAE	327
Architektura VAE	327
Analiza VAE	329
Zbieranie danych do wyszkolenia MDN-RNN	331
Szkolenie MDN-RNN	331
Architektura MDN-RNN	331
Próbkowanie z MDN-RNN	332
Szkolenie sterownika	333
Architektura sterownika	333
CMA-ES	334
Zrównoleglenie CMA-ES	336
Szkolenie In-Dream	338
Podsumowanie	341
13. Modele multimodalne	343
Wprowadzenie	344
DALL.E 2	345
Architektura	346
Koder tekstu	346
CLIP	346
Model prior	351
Dekoder	353
Przykłady z DALL.E 2	356
Imagen	360
Architektura	360
DrawBench	361

Przykłady z Imagen	362
Stable Diffusion	363
Architektura	363
Przykłady ze Stable Diffusion.	364
Flamingo	364
Architektura	365
Vision Encoder.	365
Perceiver Resampler	366
Language Model.	368
Przykłady z Flamingo	371
Podsumowanie	372
14. Wnioski	375
Oś czasu generatywnej sztucznej inteligencji.	376
2014–2017: Era VAE i GAN.	378
2018–2019: Era transformera.	378
2020–2022: Era dużego modelu.	379
Obecny stan generatywnej sztucznej inteligencji	380
Duże modele językowe	380
Modele generowania kodu na podstawie tekstu	384
Modele generowania obrazów na podstawie tekstu	385
Inne zastosowania	389
Przyszłość generatywnej sztucznej inteligencji	390
Generatywna sztuczna inteligencja w życiu codziennym	391
Generatywna sztuczna inteligencja w miejscu pracy	392
Generatywna sztuczna inteligencja w edukacji	394
Etyka i wyzwania generatywnej sztucznej inteligencji	395
Końcowe przemyślenia	397
Indeks	399
O autorze	411