18th of February, 2020

Generative Models in e-Commerce

Urs Bergmann

Zalando Research & HLEG AI, European Commission





ZALANDO - FASHION E-COMMERCE AT A EUROPEAN SCALE





helf ipython	IUsers (spasses	
continuing anyway.		-
, continuing anyway.	He GP10.sethemilags(Fe)	
	Fight Characteristic Control and Control a	
	The article and the processing of the second	
use, continuing any	y. Use GPLD. setwardings (Fe)	
in use, continuing a	way. Use GPID.sethernlags[fail	
in use, continuing	tyway. Use GPID.setvarainegi (Pa)	
	return y and the and the program is the advancement of the second	
	The critic seturation [1]	
easiy in use, contin	ng anywey' the colors set to an ing (Fa)	
tready in use, cont	aling intrastriction in the second se	
stready in use, co	anity span.	
	servery. Use GP10.setvarnings(Fal	
to stready in use	d a a a a a a a a a a a a a a a a a a a	
el la sireste la s	use Grio.setvarnings/fal al a2 al	-
enti in sirenty in	are continuing anyone to be a set of the set	
		~
The second		and the
		*0 -

Deep Learning



AN EXEMPLARY LEARNING PROBLEM - FASHION MNIST



Supervised Learning Task

- learn on pairs of images and classes $\{(x_i, y_i)\}_{i \in D}$
- predict unknown class for novel image



Fashion MNIST [Xiao et al., 2017]

MNIST [LeCun et al., 1998]

Proven Best Solutions: Deep Learning

- Define architecture
- Define objective / loss function

Optimize loss on data

Generative Modeling: a difficulty with hand-engineered losses

• hand-engineered losses are problematic, e.g.

$$MSE = \sum_{i} (f_i(x) - y_i)^2 = -\log \mathcal{L}$$

• MSE corresponds to isotropic Gaussian

$$\mathcal{L} \propto \exp\left(-\sum_{i} \left(f_i(x) - y_i\right)^2\right)$$

- assumes independence and optimizes for mean!
 - hence results in blurry images!

Possible Improvements

- VAE [Kingma&Welling, Rezende et al., 2014]
- AR models [Bengio&Bengio, 2000]
- Flows [Rezende&Mohamed, 2015]
- GANs [Goodfellow et al., 2014]





(a) Input context



(c) Context Encoder (L2 loss)

Generative Adversarial Networks [Goodfellow et al., 2014]

Core Idea

- simple noise prior p(z)
- target: transform prior to to data distribution $p(G(z)) \sim p(x)$
- solved via two-player game:
 - G: generate data from distribution
 - D: estimate probability of generated vs. real data



Original Training Value Function

$$\min_{G} \max_{D} V(D, G) = \mathbb{E}_{z \sim p_z(z)} \left[\log \left(1 - D \left(G \left(z \right) \right) \right) \right]$$
$$= \mathbb{E}_{x' \sim p_r(x)} \left[\log D(x') \right]$$

Generative Adversarial Networks



Advantages

- no need for loss in image space
- state-of-the-art image sampling/modeling

Difficulties & Limitations

- implicit model no likelihood & inference
- training hard & unclear best objective
- mode-collapse
- no fine-grained control over output
- fixed & limited output dimensionality



Selected Improvements

- DCGAN [Radford et al., 2016]
- WGAN [Arjovski et al., 2017]
- WGAN-GP [Gulrajani et al., 2017]
- SAGAN [Zhang et al., 2018]
- SN-GAN [Miyato et al., 2018]
- StyleGAN [Karras et al., 2018]
- BIGGAN [Brock et al., 2019]
- StyleGAN2 [Karras et al., 2019]

[Karras et al., 2019]

Urs Bergmann | Zalando Research

~	ALL ALL ADDRESS OF ADDRES
helf ipython	(Users/apasses
continuing anyway.	
, continuing anyway.	Are GP10.sethermlags(fp)
use, continuing any	y. Use GP10. setbarnlegs[Fail
in use, continuing a	ay. Use GP10.setwaralags(Fa)
in use, continuing	typely. Use GPID.setverslags(Fe) ortical of a law man a setup of the s
	File "Performantian in a file and a second s
ready in use, contin	ng anyway. Use GP10.setwarnings[Fa]
freedy in use, cont	alog anyway. Use GPID. setuaralogs(Fal
stready in use, co	hauten anyway. Use GPID. second market
	DU DUBIEN
t in strendy in se	east landing anyway. Use GP10.setvamings[Fal
all in strends in a	continuing anyony. Use determinings (Fall bit
static in sicressing in	any continuing support. We is the second secon
	C S L L S S S S S S S S S S S S S S S S

Fashion Design



Disentangling Input Conditionals

[Yildirim et al., 2018]

Controls & Attributes

- С Color
 - 3-dim, RGB 0
- Texture (local structure) t •
 - 512-dim 0
- \mathbf{S} Shape (mask)
 - embedded into 512-dim 0

Discriminator Loss



Real

or Fake?

Generator

Losses

 $\mathbf{x} = G(\mathbf{c}, \mathbf{t}, \mathbf{s})$

Discriminator

Generator

Color

Estimation

Color (3-dim)

Real Images

(128x128)

~	ARE ALL ADDRESS OF THE AREA
helf ipython	(Users / spasses
continuing anyway.)	
, continuing anyway.	he GP10.setwarslops(fp)
	File "Meen internation on the second design of the second se
	Contrast and the providence of
use, continuing any	y. Use GP10.sethernings(Fe)
in use, continuing a	ay. Use GPID.setvarsLogs(Fal
in use, continuing	yway. Use GPID.setvamingsffal
	white and any second state of the second state
	The ortical day's loss links to make an and links to immunity billing
wanty in use, contin	ng anyady. Use gr10.setvarnings[Fa1
iready in use, cont	alog anyway. Use GPTD.setwarnings[Fal
-t in use, CP	unuling anyway. Use GPID.setwarnings[Fal
Strend or	The state of the second s
	DÜ HIN LAN
	santing anyway. Use GP10.setvarnings[Fal
t to stream in the	at a a a a a a a a a a a a a a a a a a
all in already in a	bee GP10.setharnings(Fal al al
staff in strends in	an continuing array to barcola >1

Learning to Try On Clothes



SWAP ARTICLES ON PEOPLE: CAGAN [Jetchev and Bergmann, 2017]





SWAP ARTICLES ON PEOPLE: CAGAN [Jetchev and Bergmann, 2017]



Implicitly Learned Segmentation



Limitation

• fine details hard to learn



Configuring Pose & Outfits [Yildirim, Jetchev, Vollgraf, Bergmann, 2019]



Urs Bergmann | Zalando Research

Configuring Pose & Outfits - Unconditional Model Results



Configuring Pose & Outfits - Conditional Model Results



	NULL CONTRACTOR
helf ipython	(Users/apasses
continuing anyway.	L MILLATERITERITER
. continuing anyway.	te GPLD. setwarslags[Fe] Pro * / heimes adv. han die in mit
	The same interaction of the same interaction of the same state of the same interaction of the same state of the same sta
	Party Teaching and the second state of the sec
use, continuing any	y. Use GPID. Settemblighters
in use, continuing a	eay. Use GPID.tervarisungstration
in use, continuing	typeny. Use GPID. Setter manageriver the file free file and the setter of the setter of the set
	The "Analytical designment is a distance of the second sec
	ng anyway. Use GPID.setwarnings(Fal
andy in one, cont	uing anyway. Use GPTD.setharnings(fa)
(ready in one, con-	wites Brivery. Use GPID.setwarnings[Fal
stready in use, co	STATE F
	DU UUD LAN
	continuing anyony. Use GP10.setvarnings[Fal
	and the grant the grant seturnings (Fall and
all in streams in .	and tending anyony. Use GPID.sethamilings[Fal
static its allowants in	
-	

Time Series Forecasting:

Sales Prediction



Starting out probabilistically: DeepAR model [Salinas et al., 2017]













Simulator runs for two articles



Starting out probabilistically: DeepAR model [Salinas et al., 2017]



Limitations

- distribution needs to be specified
- time-series modeled independently
 - no article interactions!



Learning a multi-variate distribution: RealNVP [Dinh et al., 2016]

Idea

• for all bijections f it holds

$$p_X(x) = p_Z(f(x)) \left| \det\left(\frac{\partial f(x)}{\partial x^T}\right) \right|$$

• affine coupling layer

$$\begin{cases} y_{1:d} &= x_{1:d} \\ y_{d+1:D} &= x_{d+1:D} \odot \exp(s(x_{1:d})) + t(x_{1:d}) \end{cases}$$

• inverse of coupling layer

$$\Leftrightarrow \begin{cases} x_{1:d} &= y_{1:d} \\ x_{d+1:D} &= (y_{d+1:D} - t(y_{1:d})) \odot \exp\left(-s(y_{1:d})\right) \end{cases}$$





ReaINVP [Dinh et al., 2016]

Why is this great?

• affine coupling layer

$$\begin{cases} y_{1:d} &= x_{1:d} \\ y_{d+1:D} &= x_{d+1:D} \odot \exp(s(x_{1:d})) + t(x_{1:d}) \end{cases}$$

• has Jacobian

$$\frac{\partial y}{\partial x^{T}} = \begin{bmatrix} \mathbb{I}_{d} & 0\\ \frac{\partial y_{d+1:D}}{\partial x_{1:d}^{T}} & \operatorname{diag}\left(\exp\left[s\left(x_{1:d}\right)\right]\right) \end{bmatrix}$$

• efficient to calculate!





Conditioned Temporal Flows [Rasul et al., 2020]



Key Idea

 combine autoregressive model (e.g. LSTM) with conditional Flow model

Advantages

- learns distribution model from data
- able to model interactions of time-series



Performance on real-world data sets

Table 1. Test set CRPS_{sum} comparison (lower is better) of models from (Salinas et al., 2019) and our models GRU-Real-NVP, GRU-MAF and Transformer-MAF. The *two* best methods are in bold where the mean and standard errors are obtained by re-running each method three times.

Data set	Vec-LSTM ind-scaling	Vec-LSTM lowrank-Copula	GP scaling	GP Copula	GRU Real-NVP	GRU MAF	Transformer MAF
Exchange	0.008 ± 0.001	$0.007 {\pm} 0.000$	0.009 ± 0.000	0.007 ± 0.000	$0.0064{\scriptstyle\pm0.001}$	$0.005 {\scriptstyle \pm 0.001}$	$0.005 {\scriptstyle \pm 0.001}$
Solar	$0.391{\scriptstyle \pm 0.017}$	$0.319{\scriptstyle \pm 0.011}$	$0.368{\scriptstyle\pm0.012}$	$0.337{\scriptstyle\pm0.024}$	$0.331{\scriptstyle \pm 0.02}$	$0.315 {\scriptstyle \pm 0.023}$	$0.301 {\scriptstyle \pm 0.014}$
Electricity	$0.025{\scriptstyle\pm0.001}$	$0.064 {\pm} 0.008$	$0.022{\scriptstyle\pm0.000}$	$0.024{\scriptstyle\pm0.002}$	0.024 ± 0.001	$0.0208 {\scriptstyle \pm 0.000}$	$0.0207 {\scriptstyle \pm 0.000}$
Traffic	$0.087 {\pm} 0.041$	$0.103{\scriptstyle \pm 0.006}$	$0.079{\scriptstyle\pm0.000}$	$0.078{\scriptstyle\pm0.002}$	0.078 ± 0.001	$0.069 {\scriptstyle \pm 0.002}$	$0.056 {\scriptstyle \pm 0.001}$
Taxi	$0.506{\scriptstyle \pm 0.005}$	$0.326{\scriptstyle\pm0.007}$	$0.183{\scriptstyle \pm 0.395}$	$0.208{\scriptstyle\pm0.183}$	$0.175 {\scriptstyle \pm 0.001}$	0.161 ± 0.002	$0.179{\scriptstyle \pm 0.002}$
Wikipedia	$0.133{\scriptstyle \pm 0.002}$	$0.241{\scriptstyle\pm0.033}$	$1.483{\scriptstyle \pm 1.034}$	$0.086{\scriptstyle \pm 0.004}$	$0.078{\scriptstyle \pm 0.001}$	0.067 ± 0.001	$0.063 {\scriptstyle \pm 0.003}$



CONCLUSION

Generative Image Modeling

- supervised losses + deep learning work excellent on many tasks
- extension of conditional GANs allow conditional image creation for real-time fashion exploration
- GANs allow generation of data even in face of lacking training data

Time-Series Modeling

• Combining Flows with AR models yields powerful multi-variate time-series models







Thanks to ...

Computer Vision

- Christian Bracher
- Sebastian Heinz
- Siavash Haghiri
- Roland Vollgraf

Image Creation

- Nikolay Jetchev
- Gokhan Yildirim

Intelligent Control

- Kashif Rasul
- Ingmar Schuster
- Saboor Sheikh
- Calvin Seward

Natural Language Processing

• Josip Krapac

THANK YOU!



We're hiring!

jobs.zalando.com

