Generative Models in e-Commerce

Urs Bergmann
Zalando Research &
HLEG AI, European Commission
<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revenue (2019)</td>
<td>~ 6.5 billion EUR</td>
</tr>
<tr>
<td>Employees in Europe</td>
<td>~ 14,000</td>
</tr>
<tr>
<td>Mobile visits (%)</td>
<td>&gt; 70%</td>
</tr>
<tr>
<td>Orders in 2019</td>
<td>~ 145 million</td>
</tr>
<tr>
<td>Active customers</td>
<td>31 million</td>
</tr>
<tr>
<td>Product choices</td>
<td>&gt; 500,000</td>
</tr>
<tr>
<td>Brands</td>
<td>&gt; 2,500</td>
</tr>
<tr>
<td>Countries</td>
<td>17</td>
</tr>
</tbody>
</table>
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

Proven Best Solutions: Deep Learning

- Define architecture
- Define objective / loss function

Optimize loss on data

MNIST [LeCun et al., 1998]

Fashion MNIST [Xiao et al., 2017]
Generative Modeling: a difficulty with hand-engineered losses

- hand-engineered losses are problematic, e.g.
  \[ \text{MSE} = \sum_i (f_i(x) - y_i)^2 = -\log \mathcal{L} \]

- MSE corresponds to isotropic Gaussian
  \[ \mathcal{L} \propto \exp \left( -\sum_i (f_i(x) - y_i)^2 \right) \]

- assumes independence and optimizes for mean!
  \[ \rightarrow \text{ hence results in blurry images!} \]

Possible Improvements
- VAE [Kingma&Welling, Rezende et al., 2014]
- AR models [Bengio&Bengio, 2000]
- Flows [Rezende&Mohamed, 2015]
- GANs [Goodfellow et al., 2014]
**Generative Adversarial Networks** [Goodfellow et al., 2014]

**Core Idea**
- simple noise prior $p(z)$
- target: transform prior 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**
\[
\min_G \max_D V(D, G) = \mathbb{E}_{z \sim p_z(z)} \left[ \log (1 - D(G(z))) \right] \\
= \mathbb{E}_{x' \sim p_x(x)} \left[ \log D(x') \right]
\]
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]
Fashion Design
Disentangling Input Conditionals
[Yildirim et al., 2018]

Controls & Attributes

- **Color** $c$
  - 3-dim, RGB
- **Texture** (local structure) $t$
  - 512-dim
- **Shape** (mask) $s$
  - embedded into 512-dim

Discriminator Loss

$$ \max_D \mathcal{L}_W - \lambda_g \mathcal{L}_{gp} - \mathcal{L}_{aux} $$

“Conditional Image Synthesis With Auxiliary Classifier GANs”
[Ódena et. al., 2017]
Learning to Try On Clothes
SWAP ARTICLES ON PEOPLE: CAGAN [Jetchev and Bergmann, 2017]

MISSING!
NO TRAINING DATA!
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]

Method
- extend StyleGan [Karras et al, 2019]
- condition on article & pose embedding + random vector

Data
Configuring Pose & Outfits - Unconditional Model Results
Configuring Pose & Outfits - Conditional Model Results

Outfit #1

Outfit #2
Time Series Forecasting:
Sales Prediction
Starting out probabilistically: DeepAR model [Salinas et al., 2017]
Starting out probabilistically: DeepAR model

Training: loss is log likelihood of observed value
Starting out probabilistically: DeepAR model

inference: sample from distribution to get quantiles etc.
Simulator runs for two articles

averaged simulations

simulation 1

simulation 2
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
  $$\begin{align*}
  y_{1:d} &= x_{1:d} \\
  y_{d+1:D} &= x_{d+1:D} \odot \exp \left( s(x_{1:d}) \right) + t(x_{1:d})
  \end{align*}$$

- inverse of coupling layer
  $$\begin{align*}
  x_{1:d} &= y_{1:d} \\
  x_{d+1:D} &= (y_{d+1:D} - t(y_{1:d})) \odot \exp \left( - s(y_{1:d}) \right)
  \end{align*}$$
RealNVP [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
  \[
  \frac{\partial y}{\partial x^T} = \begin{bmatrix}
  I_d \\
  \frac{\partial y_{d+1:D}}{\partial x_{1:d}^T} \text{diag}(\exp(s(x_{1:d})))
  \end{bmatrix}
  \]
- 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.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Vec-LSTM ind-scaling</th>
<th>Vec-LSTM lowrank-Copula</th>
<th>GP scaling</th>
<th>GP Copula</th>
<th>GRU Real-NVP</th>
<th>GRU MAF</th>
<th>Transformer MAF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exchange</td>
<td>0.008±0.001</td>
<td>0.007±0.000</td>
<td>0.009±0.000</td>
<td>0.007±0.000</td>
<td><strong>0.0064±0.001</strong></td>
<td>0.005±0.001</td>
<td><strong>0.005±0.001</strong></td>
</tr>
<tr>
<td>Solar</td>
<td>0.391±0.017</td>
<td>0.319±0.011</td>
<td>0.368±0.012</td>
<td>0.337±0.024</td>
<td>0.331±0.02</td>
<td>0.315±0.023</td>
<td><strong>0.301±0.014</strong></td>
</tr>
<tr>
<td>Electricity</td>
<td>0.025±0.001</td>
<td>0.064±0.008</td>
<td>0.022±0.000</td>
<td>0.024±0.002</td>
<td>0.024±0.001</td>
<td>0.0208±0.000</td>
<td><strong>0.0207±0.000</strong></td>
</tr>
<tr>
<td>Traffic</td>
<td>0.087±0.041</td>
<td>0.103±0.006</td>
<td>0.079±0.000</td>
<td>0.078±0.002</td>
<td>0.078±0.001</td>
<td>0.069±0.002</td>
<td><strong>0.056±0.001</strong></td>
</tr>
<tr>
<td>Taxi</td>
<td>0.506±0.005</td>
<td>0.326±0.007</td>
<td>0.183±0.395</td>
<td>0.208±0.183</td>
<td><strong>0.175±0.001</strong></td>
<td>0.161±0.002</td>
<td>0.179±0.002</td>
</tr>
<tr>
<td>Wikipedia</td>
<td>0.133±0.002</td>
<td>0.241±0.033</td>
<td>1.483±1.034</td>
<td>0.086±0.004</td>
<td>0.078±0.001</td>
<td><strong>0.067±0.001</strong></td>
<td><strong>0.063±0.003</strong></td>
</tr>
</tbody>
</table>
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

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

Image Creation
- Nikolay Jetchev
- Gokhan Yildirim

Natural Language Processing
- Josip Krapac

THANK YOU!
We’re hiring!

jobs.zalando.com