Build a Recommendation Engine on AWS
Today

Yotam Yarden
Data Scientist, Amazon Web Services
Agenda

• **Recommendation Engine – Why?**

• Recommendation Engine – Common Techniques

• Introducing Amazon SageMaker

• Develop, Train & Deploy a Recommendation Engine in 15 minutes

• Customer use cases
Welcome to Amazon.com Books!

One million titles, consistently low prices.

(If you explore just one thing, make it our personal notification service. We think it's very cool!)

SPOTLIGHT! -- AUGUST 16TH
These are the books we love, offered at Amazon.com low prices. The spotlight moves EVERY day so please come often.

ONE MILLION TITLES
Search Amazon.com's million title catalog by author, subject, title, keyword, and more... Or take a look at the books we recommend in over 20 categories... Check out our customer reviews and the award winners from the Hugo and Nebula to the Pulitzer and Nobel... and bestsellers are 30% off the publishers list...

EYES & EDITORS, A PERSONAL NOTIFICATION SERVICE
Like to know when that book you want comes out in paperback or when your favorite author releases a new title? Eyes, our tireless, automated search agent, will send you mail. Meanwhile, our human editors are busy previewing galleys and reading advance reviews. They can let you know when especially wonderful works are published in particular genres or subject areas. Come in, meet Eyes, and have it all explained.

YOUR ACCOUNT
Check the status of your orders or change the email address and password you have on file with us. Please note that you do not need an account to use the store. The first time you place an order, you will be given the opportunity to create an account.
And today…
Motivation

- Personalize and enhance customer experience

- Different goals:
  - Increased time spent on a platform
  - Suggest complementary items
  - Customer satisfaction
Use Cases

Ecommerce:
  • Amazon.com

Content:
  • Movies (Netflix)
  • Music (Amazon Music)
  • Articles (The Global And Mail)

Finance:
  • Services Recommendation
  • Stocks buying / selling
  • Relevant news and stock related data

Education:
  • Courses recommendations

Legal:
  • Similar cases
Agenda

• Recommendation Engine – Why?
• Recommendation Engine – Common Techniques
• Introducing Amazon SageMaker
• Develop, Train & Deploy a Recommendation Engine in 15 minutes
• Customer Use Cases
Test / Validation
Naïve approach

Linear model? [type of user, movie genre, etc.]

Polynomial model? [+interactions]
Matrix Factorization – “Neural Networks” Representation
Deep Matrix Factorization
Binary Predictions
Binary Predictions

+Negative Sampling
Most of the Data is Still Untapped

- Images
- Titles
- Descriptions
- Reviews
- Episode Names
DSSM – Deep Structures Semantic Models

User Search
BOW

Embedding
User

Item
title words
resnet: imgs

dropout
BOW
Embedding

dense
concat

dense

output

score
## Which Technique to Choose? Roadmap Matrix

<table>
<thead>
<tr>
<th>Iterative process</th>
<th>Data Available</th>
<th>Relevant Algorithms</th>
<th>Relative Complexity</th>
<th>Deployment Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Historical data size – 30d / 60d / 1y…</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Fine-tuning techniques (daily, weekly..)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Inference - compressed model? Tradeoff between model complexity and inference latency</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Validation system setup</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Iterate fast and simple</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Limited user data</th>
<th>User data</th>
<th>More user data</th>
<th>Extensive user data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary user-item interaction</td>
<td>User data</td>
<td>Additional user-item interaction</td>
<td>More user data</td>
<td>Extensive user data</td>
</tr>
<tr>
<td>Matrix Factorization</td>
<td>Matrix Factorization</td>
<td>Factorization Machines DiFacto</td>
<td>DSSM</td>
<td>Customized and more advanced DSSM</td>
</tr>
</tbody>
</table>

|                | 2 | 4 | 5 | 5 |

© 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Agenda

• Recommendation Engine – Why?
• Recommendation Engine – Common Techniques
• Introducing Amazon SageMaker
• Develop, Train & Deploy a Recommendation Engine in 15 minutes
• Customer Use Cases
ML @ AWS: Our mission

Put machine learning in the hands of every developer and data scientist
Customer Running ML on AWS Today
ML is still too complicated for everyday developers and data scientists

- Collect and prepare training data
- Choose and optimize your ML algorithm
- Set up and manage environments for training
- Train and tune model (trial and error)
- Deploy model in production
- Scale and manage the production environment
Amazon SageMaker

Easily build, train, and deploy machine learning models
Amazon SageMaker

Pre-built notebooks for common problems

BUILD

Choose and optimize your ML algorithm
Set up and manage environments for training
Train and tune model (trial and error)
Deploy model in production
Scale and manage the production environment
Amazon SageMaker

**Pre-built notebooks for common problems**

**Built-in, high performance algorithms**

**ALGORITHMS**
- K-Means Clustering
- Principal Component Analysis
- Neural Topic Modelling
- Factorization Machines
- Linear Learner - Regression
- XGBoost
- Latent Dirichlet Allocation
- Image Classification
- Seq2Seq
- Linear Learner - Classification

**FRAMEWORKS**
- Apache MXNet
- TensorFlow
- Caffe2, CNTK, PyTorch, Torch

**BUILD**

Set up and manage environments for training
Train and tune model (trial and error)
Deploy model in production
Scale and manage the production environment

© 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Amazon SageMaker

**Build**
- Pre-built notebooks for common problems
- Built-in, high performance algorithms

**Train**
- One-click training

- Train and tune model (trial and error)
- Deploy model in production
- Scale and manage the production environment

© 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Amazon SageMaker

Pre-built notebooks for common problems

Built-in, high performance algorithms

One-click training

Hyperparameter optimization

Deploy model in production

Scale and manage the production environment

BUILD

TRAIN
Amazon SageMaker

**BUILD**
- Pre-built notebooks for common problems

**TRAIN**
- One-click training
- Hyperparameter optimization

**DEPLOY**
- One-click deployment

Scale and manage the production environment
Amazon SageMaker

- Pre-built notebooks for common problems
- Built-in, high performance algorithms
- One-click training
- Hyperparameter optimization
- One-click deployment
- Fully managed hosting with auto-scaling
Agenda

• Recommendation Engine – Why?
• Recommendation Engine – Common Techniques
• Introducing Amazon SageMaker
• Develop, Train & Deploy a Recommendation Engine in 15 minutes
• Customer Use Cases
console
Agenda

• Recommendation Engine – Why?

• Recommendation Engine – Common Techniques

• Introducing Amazon SageMaker

• Develop, Train & Deploy a Recommendation Engine in 15 minutes

• Customer Use Cases
Customers Use Cases

“Erento’s in-house Data Science team is using Amazon SageMaker to build and deploy ML models to solve item availability and decrease the enquiry-to-offer time through a recommendation system, which suggests similar items that are available and increases the chance for a successful booking. Using Amazon SageMaker reduced our recommendation system building time from half a year to few weeks and reduced the algorithm training time from hours to few seconds. It also helped us reduce dependencies between projects, which has streamlined our whole pre-deployment process.

- Wassim Zoghlami, Data Scientist Engineer at Erento

“Using machine learning, we can provide better recommendations for our clients and enhance their customer experience. The AWS ML Acceleration Program delivered by the Professional Services Team, was really useful and suited our business needs. We believe that with Amazon SageMaker we can build a great recommendation system, and will be able to scale our ML training and deployment jobs in a more simple and faster way.

- Igor Veremchuk - Director of Engineering at Datajet

“Once we at HolidayPirates decided to take a strategic step towards personalization, we wanted to move fast. With the help of AWS Professional Services and the account team introducing us to Amazon SageMaker we are now able to develop, train and deploy recommendation system models in a very short time and independently from any other department. We no longer need to wear the hats of IT, big data, data science etc, and we can focus on what is important for our customers and enhance their user experience.

- Bojan Kostic, Data Team Lead at HolidayPirates
References

- https://github.com/apache/incubator-mxnet
- https://github.com/awslabs/amazon-sagemaker-examples
- https://www.youtube.com/watch?v=cftJAwKWeA
- https://www.youtube.com/watch?v=1cRGpDXtJC8&t=640s
GO BUILD