aws machine learning

End-to-end natural language processing with Amazon SageMaker

Julien Simon
Global Evangelist, AI & Machine Learning, AWS
@julsimon

Challenge #1: preparing data

- NLP datasets are often very large: millions of sentences, emails, etc.
- They need extensive processing
  - Cleaning: removing punctuation, numbers, words that don't add context (aka stop words).
  - Transformation: bag of words, word vectors
  - Conversion: input format expected by your algorithm
- … again and again, as this is a highly iterative process
- This requires tools and infrastructure, and maintaining them takes valuable time away from the actual ML work

Challenge #2: selecting algorithms

- NLP problems are extremely diverse and require a wide range of algorithms
  - Off the shelf algorithms
  - Pre-trained models available in model zoos (TF Hub, PyTorch Hub, etc.)
  - Your own algorithms

- We need a standardized environment that makes it easy to experiment
  - Same workflow, same IDE, same SDK
  - Tracking and comparing trials
  - Visualizing results

Challenge #3: training and deploying

- Even a simple project will require hundreds of training jobs
- Infrastructure capacity and cost can limit speed of iteration
- Different algorithms require different infrastructure
- Infrastructure is just the beginning: what about tuning, debugging, monitoring, scaling, and so on?
- Again, maintaining this yourself takes valuable time away from your actual ML work

The AWS ML Stack

Broadest and most complete set of Machine Learning capabilities

AI SERVICES

VISION	SPEECH		TEXT			SEARCH	CHATBOTS	PERSONALIZATION	FORECASTING	FRAUD	DEVELOPMENT	CONTACT CENTERS
Amazon Rekognition	Amazon Polly	Amazon Transcribe +Medical	Amazon Comprehend +Medical	Amazon Translate	Amazon Textract	Amazon Kendra	Amazon Lex	Amazon Personalize	Amazon Forecast	Amazon Fraud Detector	Amazon CodeGuru	Contact Lens For Amazon Connect

ML SERVICES

Amazon SageMaker

Ground Truth	ML Marketplace	SageMaker Studio IDE								Neo	A2I
		Built-in algorithms	Notebooks	Experiments	Processing & Model Evaluation	Model training & tuning	Debugger	Autopilot	Model hosting	Model Monitor	

ML FRAMEWORKS & INFRASTRUCTURE

TensorFlow, mxnet, PYTORCH, GLUON, Keras, DeepGraphLibrary

Deep LearningAMIs & Containers	GPUs &CPUs	ElasticInference	Inferentia	FPGA

aws machine learning

https://ml.aws

Amazon SageMaker helps you build, train, and deploy models

Prepare | Build | Train & Tune | Deploy & Manage

Web-based IDE for machine learning

Automatically build and train models

Fully managed data processing jobs and data labeling workflows

Collect and prepare training data

One-click collaborative notebooks and built-in, high performance algorithms and models

Choose or build an ML algorithm

One-click training

Set up and manage environments for training

Debugging and optimization

Train, debug, and tune models

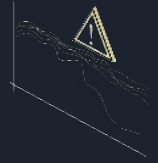
Visually track and compare experiments

Manage training runs

One-click deployment and autoscaling

Deploy model in production

Automatically spot data drift



Monitor models

Add human review of predictions

Validate predictions

Fully managed with auto-scaling for 75% less

Scale and manage the production environment

Same service and APIs, from experimentation to production

Model options

AWS Marketplace for Machine Learning

Training code

Amazon SageMaker AutoPilot

Factorization Machines
Linear Learner
Principal Component Analysis
K-Means Clustering
XGBoost

Built-in Algorithms (17)
No ML coding required

mxnet Chainer TensorFlow PyTorch scikit-learn dmlc XGBoost

Built-in Frameworks
Bring your own code
Use open source containers

Bring Your Own
Full control, run your container
R, C++, etc.

Fully managed training, spot instances included

Demo: topic modeling on a million news headlines

Data preparation with Amazon SageMaker Processing

Model training and deployment with Amazon SageMaker, using the Latent Dirichlet Allocation (LDA) and Neural Topic Modeling (NTM) built-in algorithms

Experiment tracking with Amazon SageMaker Experiments

https://ml.aws

https://aws.amazon.com/sagemaker
https://github.com/aws/sagemaker-python-sdk
https://github.com/awslabs/amazon-sagemaker-examples

https://youtube.com/juliensimonfr
https://medium.com/@julsimon

https://bit.ly/3lS6n7o

https://github.com/PacktPublishing/Learn-Amazon-SageMaker

aws machine learning