Data Science Indy: News

July 2019

Data Science News

Google announced the availability of preconfigured container images with deep learning frameworks and tools.
https://www.forbes.com/sites/janakirammsv/2019/07/01/google-just-made-machine-learning-more-accessible-and-portable-with-containers/#6ad4b0894dcc

Facebook AI Research announced the open-source release of a deep-learning recommendation model, DLRM, that achieves state-of-the-art accuracy in generating personalized recommendations.
https://www.infoq.com/news/2019/07/facebook-recommendation-model/

Tensorflow has released a pruning API:
https://medium.com/tensorflow/tensorflow-model-optimization-toolkit-pruning-api-42cac9157a6a

This is in addition to other work Google has been doing on improving performance, cost to serve, and carbon footprint
https://ai.googleblog.com/2019/05/efficientnet-improving-accuracy-and.html?m=1

Data Science News

"Computer science curriculums should spend more time emphasizing how deep learning works and especially the limitations of today's correlative pattern encoders."
https://www.forbes.com/sites/kalevleetaru/2019/07/10/why-is-there-so-little-understanding-of-deep-learnings-limitations/#a3c3f97b96fd

Contrary to McKinsey and others stating that AI may kill or displace millions of workers, ZipRecruiter asserted AI created about three times as many jobs as it took away in 2018
https://www.fastcompany.com/90369739/is-ai-killing-jobs-actually-it-added-3x-more-in-2018

Data Science News

Author pulls program which attempts to create inappropriate photos of unwitting image subjects
https://arstechnica.com/tech-policy/2019/06/author-pulls-software-that-used-deep-learning-to-virtually-undress-women/

Deep learning has a terrible carbon foot print: a single AI model can emit as much carbon as five cars in their lifetimes
http://arxiv.org/abs/1906.02243

Facebook stole Bill Gate's voice
https://www.theverge.com/2019/6/10/18659897/ai-voice-clone-bill-gates-facebook-melnet-speech-generation

Links or Discussion

Notebook with several top Kaggle implementations:
https://www.kaggle.com/sudalairajkumar/winning-solutions-of-kaggle-competitions

PyTorch has its own Model Zoo: https://pytorch.org/hub

AWS offers one competing approach to data quality:
https://aws.amazon.com/blogs/big-data/test-data-quality-at-scale-with-deequ/

Word embeddings may not be as biased as once thought:
https://twitter.com/rikvannoord/status/1132933236756238341