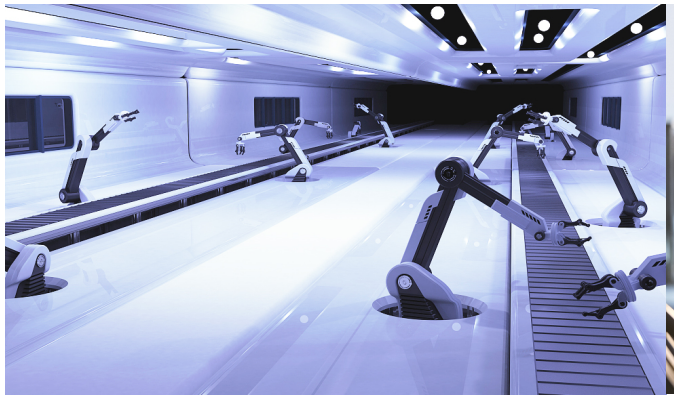
New Generation of Deep Learning Network: Auto Deep Learning and its Applications

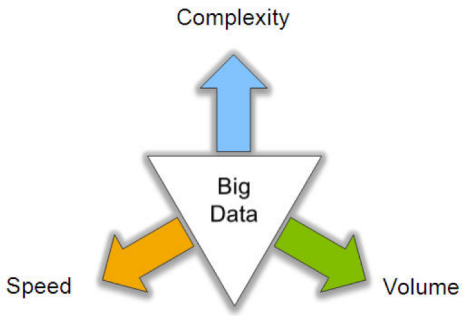
Dejing Dou
Head of Big Data Lab (BDL), Baidu Research

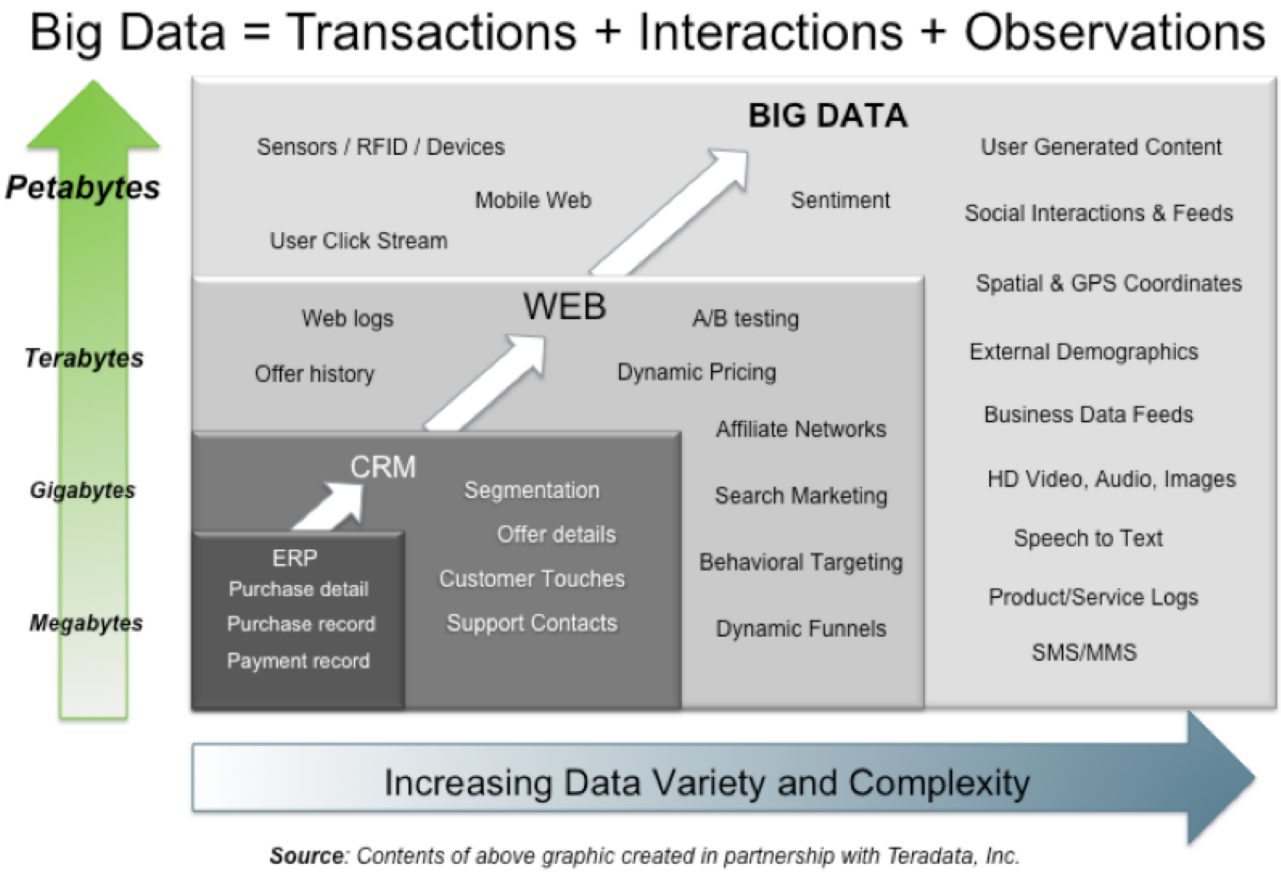


Big Data and Artificial Intelligence will become the “engines” of our times!



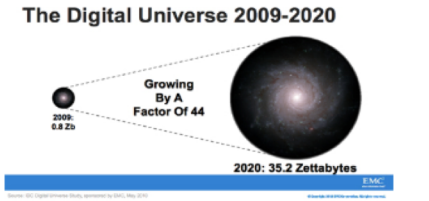
Big Data: 3V's

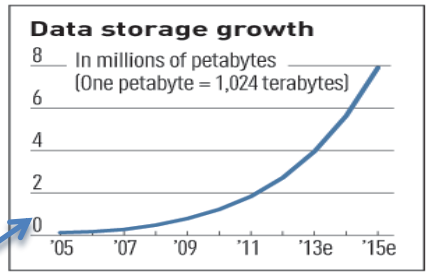


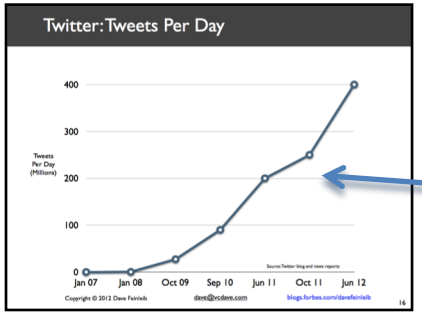


***Source**: Contents of above graphic created in partnership with Teradata, Inc.*

- **Data Volume**
  - 44x increase from 2009 to 2020
  - From 0.8 zettabytes to 35zb
- Data volume is increasing exponentially



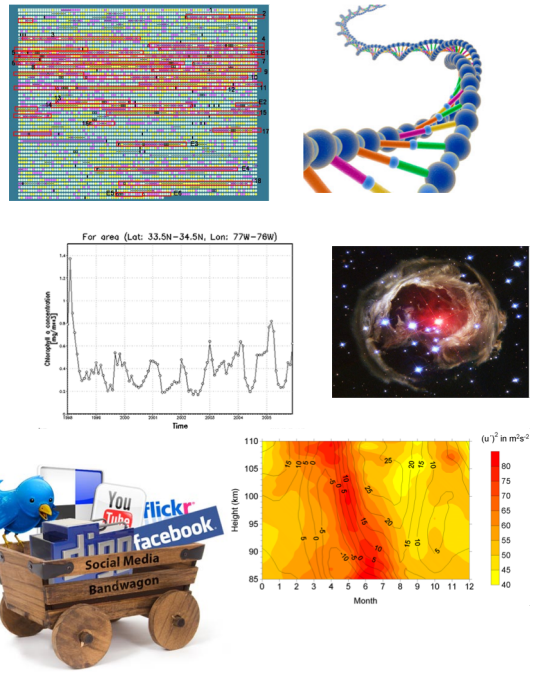




Variety (Complexity)

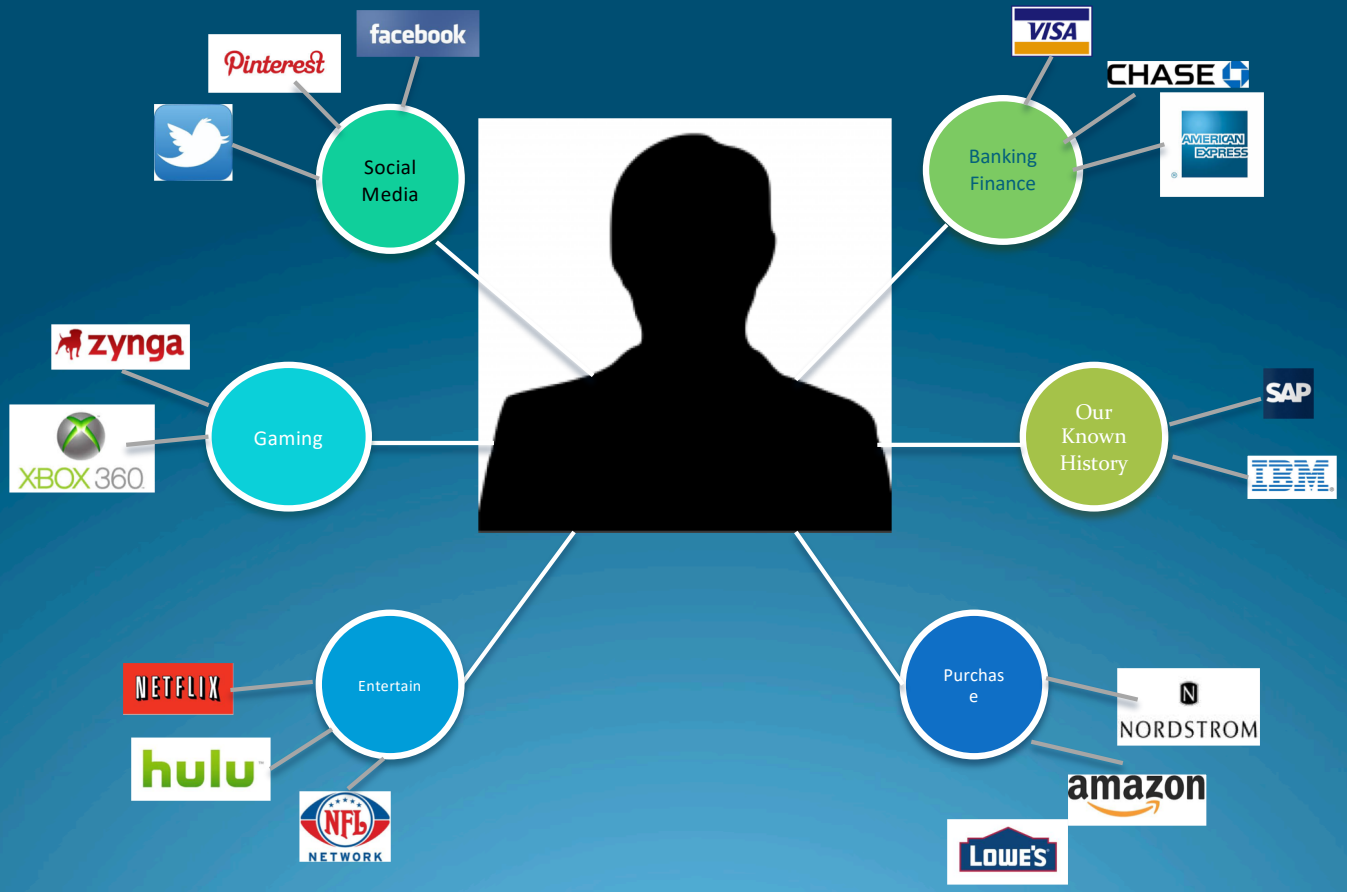
- Relational Data (Tables/Transaction/Legacy Data)
- Text Data (Web, HTML)
- Semi-structured Data (XML)
- Graph Data
  - Social Network, Semantic Web (RDF, LODs), …
- Streaming Data
  - You can only scan the data once
- A single application can be generating/collecting many types of data
- Big Public Data (online, weather, finance, etc)

Heterogeneous in syntax and semantics in general



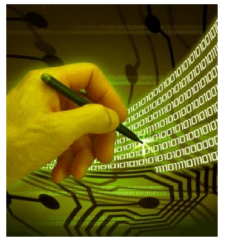
To extract knowledge➔ all these types of data need to linked/integrated

A Single View to the Customer



Velocity (Speed)

- Data is generated fast and need to be processed fast (with some direction)
- Online Data Analytics
- Late decisions ➔ missing opportunities
- **Examples**
  - **E-Promotions:** Based on your current location, your purchase history, what you like ➔ send promotions right now for store next to you
  - **Healthcare monitoring:** sensors monitoring your activities and body ➔ any abnormal measurements require immediate reaction
  - **Credit card fraud detection**
  - **Gravity Wave Detection**

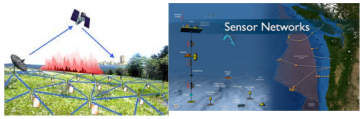


Real-time/Fast Data

(all of us are generating data)

(collecting all sorts of data)

(tracking all objects all the time)



(measuring all kinds of data)

- The progress and innovation is no longer hindered by the ability to collect data
- But, by the ability to manage, analyze, summarize, visualize, and discover knowledge from the collected data in a timely manner and in a scalable fashion

About Big Data Lab, Baidu Research

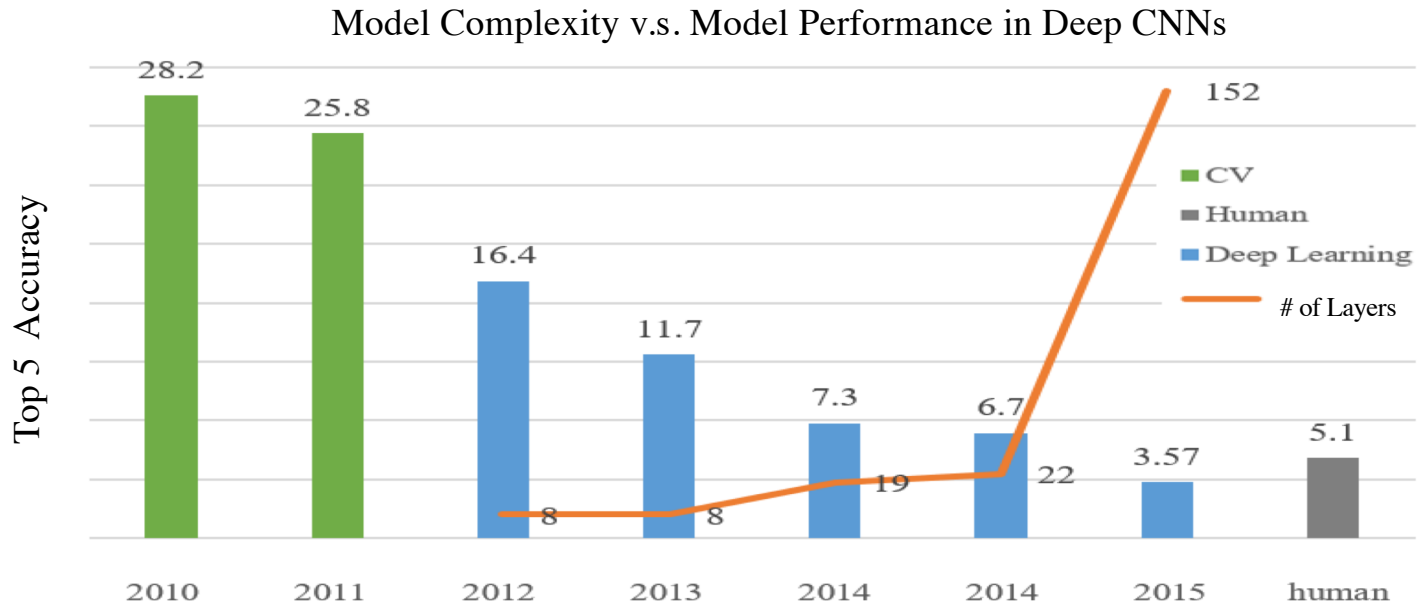
- Big Data Lab was founded in collaboration with the United Nations in 2014

- Our development consists of two main stages:
  - 2014-2017: How to build models when we have access to big data
  - 2018-: How to enable enterprises/individuals to build and explain models even if they do not have access to big data
    - → "Everyone can build their own models"
    - → Ongoing: "Everyone can explain their own models"

- Our work spans multiple research areas
  - Neural Architecture Search (NAS)
  - Transfer Learning
  - Meta-Learning and Federated Learning
  - Data Augmentation
  - Interpretability of Deep Neural Network

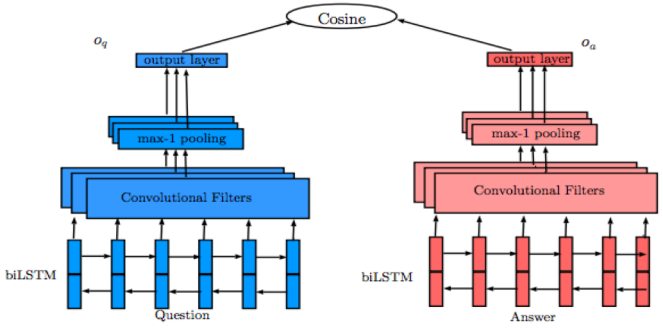
Big Data, Deep Learning and AI Driven by Big Computing Power

- Deep Learning becomes the core power of AI
- It is almost a definite trend that AI applications will be industrialized, standardized, modularized and automated.

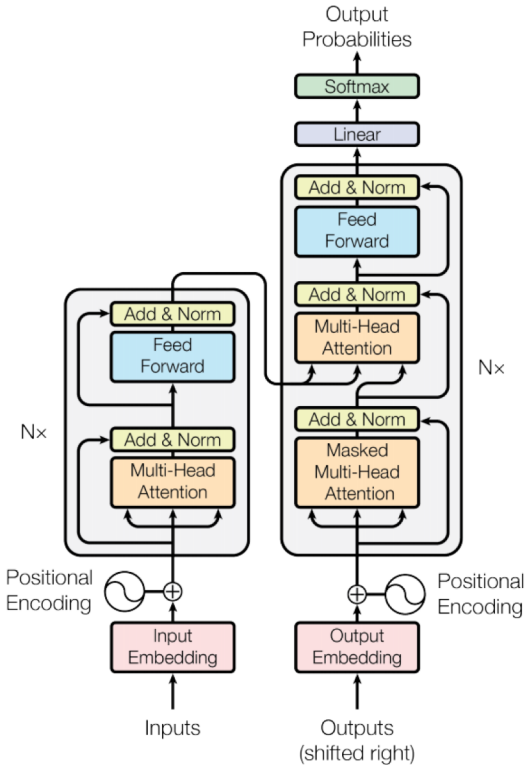


Application of Deep Learning in Natural Language Processing

- Deep Learning is widely applicable in NLP tasks
  - Information Retrieval
    - Named Entity Recognition
    - Part-of-speech Tagging
    - Entity relation extraction
  - Text Classification
  - Sentimental Analysis
  - Semantic Analysis
  - Q&A
  - Language Generation
  - Machine Translation
  - ... ...



Q&A model based on CNN and LSTM

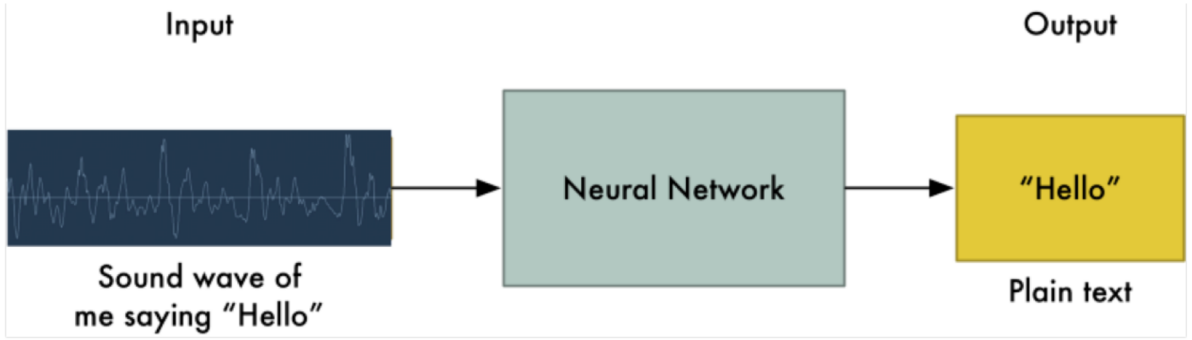


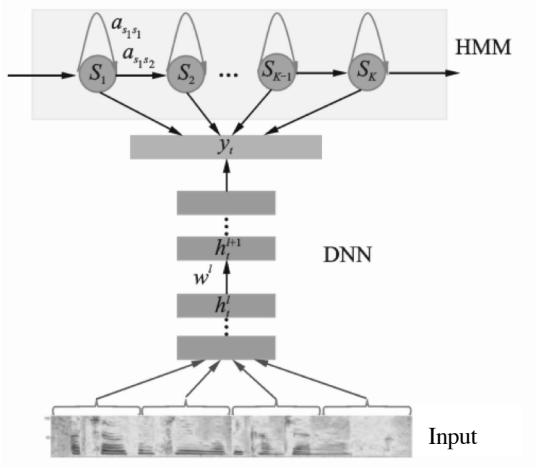
Transfer Learning model based on Attention

Otter D W, Medina J R. A survey of the usages of deep learning in natural language processing

Application of Deep Learning in Speech Recognition

- Deep Learning is the main method in speech recognition





Speech Recognition System
based on DNN

AI Customer Service

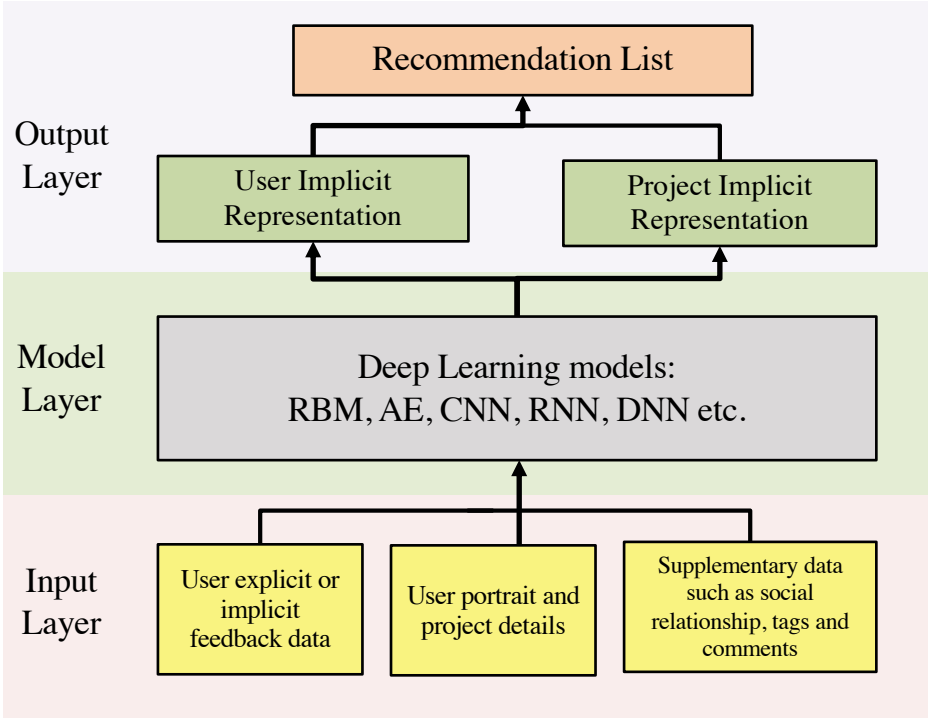
Voice Input

Voice Assistant

Vadwala A Y, Suthar K A, Karmakar Y A, et al. Survey paper on Different Speech Recognition Algorithm: Challenges and Techniques[J]. Int. J. Comput. Appl., 2017, 175(1): 31-36.

Application of Deep Learning in Recommender System

- Deep Learning is largely advantageous in Recommender System
  - Directly extracts features from data: Expressive
  - Easy to process noisy data: Robust
  - RNN can handle series data
  - Can more accurately learn user/item features
  - Facilitates standardized data processing



Recommender System based on Deep Learning

Zhang S, et al. Deep learning based recommender system: A survey and new perspectives[J].

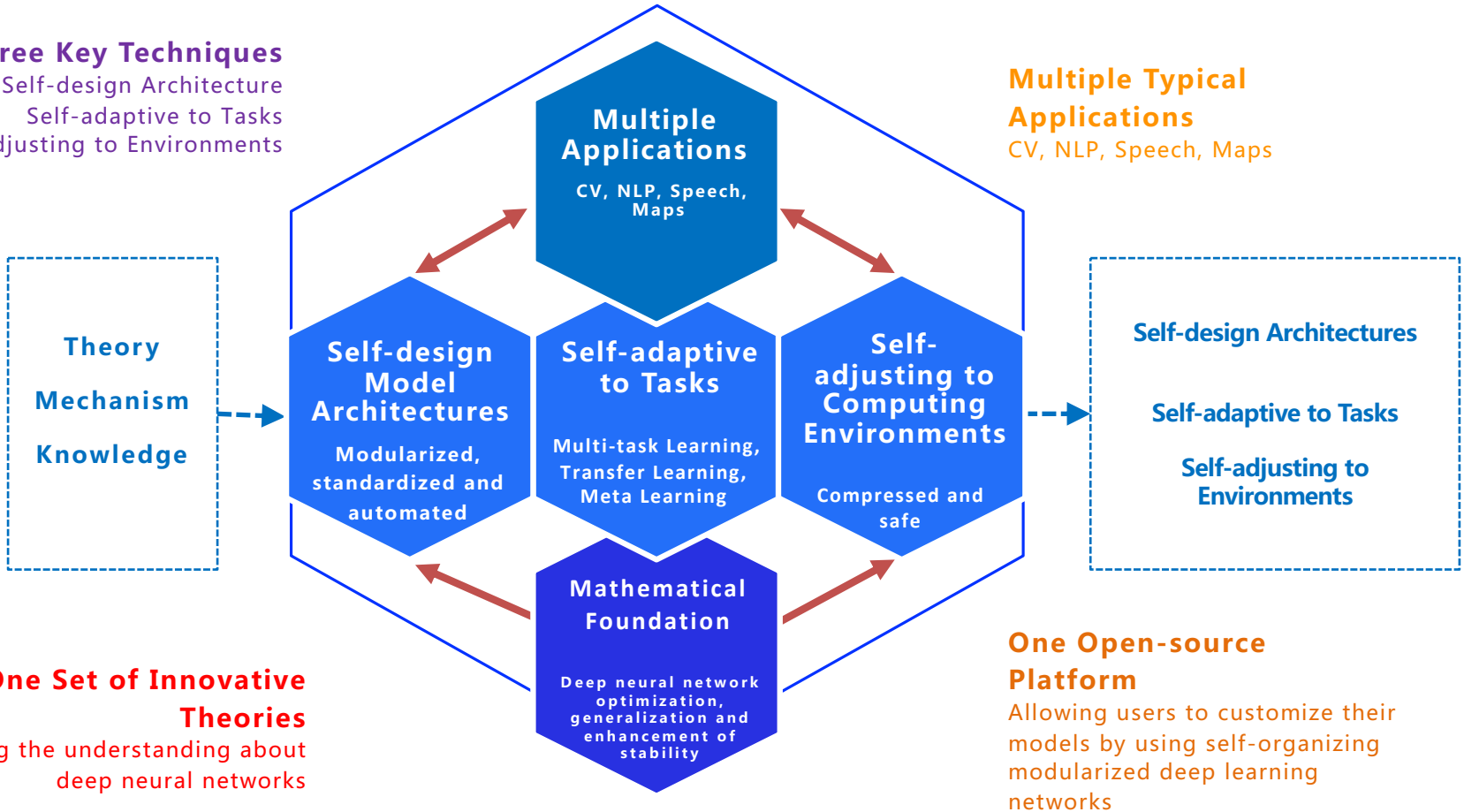
Current Limitations in Deep Learning

- Deep Learning needs to be scalable, automated and self-adapting to various scenarios/applications/hardware/modalities
- Current: Designed by a small group of professional researchers/engineers; time-consuming to fine-tune the hyper-parameters
- Future: Scalable, automated and customized design
- Current: Commonly deployed on servers or in the clouds
- Future: Ubiquitously deployed on low-cost and heterogenous computing hardware ends
- Current: Designed for a particular scenario, largely dependent on data
- Future: Self-organizing, self-adaptive, self-evolving; reduced dependence on data
- Current: Blackbox solution that is hard to interpret; prone to adversarial attacks
- Future: Enhanced interpretability; robust to adversarial attacks

Key Techniques in AutoDL

Three Key Techniques
Self-design Architecture
Self-adaptive to Tasks
Self-adjusting to Environments

Multiple Typical Applications
CV, NLP, Speech, Maps



One Set of Innovative Theories
Deepening the understanding about deep neural networks

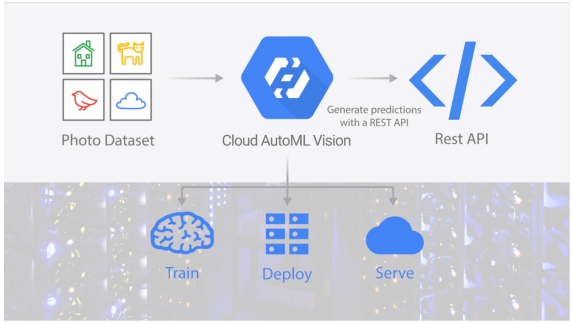
One Open-source Platform
Allowing users to customize their models by using self-organizing modularized deep learning networks

Global Trend of Auto Deep Learning

- Google: Cloud AutoML and DARTS (Differential Architecture Search)
- Microsoft: MS Custom Vision, MS Azure ML
- Amazon: Amazon ML
- Salesforce: Transmogrif AI

Google Cloud AutoML

- Cloud AutoML is a suite of machine learning products that enables developers with limited machine learning expertise to train high-quality models specific to their business needs.
- Interface: Interactive graphical user interface
- Core Algorithms: Transfer Learning, Learning2learn, NAS.
- Service: Translation, NLP, CV etc.
- Data Preparation:
  - Needs to provide input images and the labels
  - AutoML Vision requires at least 100 images for each class
  - Visualization tools to visualize data
- Application
  - Education, healthcare, finance, games, media & entertainment, retail, government.



Microsoft Custom Vision

- Start training a customized computer vision model by simply uploading and labeling a few images with labels. Custom Vision Service will do the rest. A single click allows a user to export the model after training. The model can execute locally on in a Docker container.
- Interface: Interactive graphical user interface
- Core Algorithm: Transfer Learning
- Service: CV
- Image Upload:
  - Upload unlabeled images, or use Custom Vision Service to quickly label the images
- Training:
  - Custom Vision Service will learn from the labeled images

Amazon ML

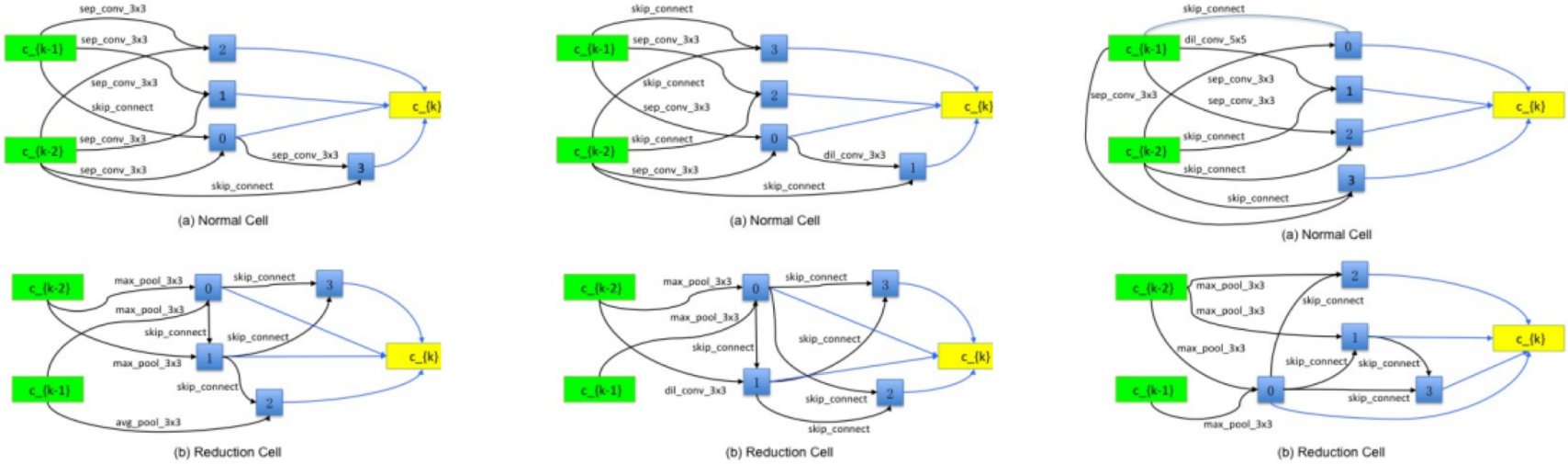
- Amazon ML is a powerful cloud service. It provides a user with the ability to build, train and deploy machine learning models quickly regardless of their technical background
- Interface: Interactive graphical user interface (Jupyter Notebook and APIs)
- Data Pre-processing:
  - This service can obtain data from multiple sources, including Amazon RDS, Amazon Redshift, and CSV files
  - It does not require data pre-processing
  - Jupyter Notebook allows easy visualization of data
- Model Training:
  - Amazon ML is capable of learning many tasks: binary classification, multi-class classification and regression
  - Amazon ML will choose the appropriate model upon seeing the input data

SalesForce Transmogrif AI

- TransmogrifAI is an end-to-end AutoML library for structured data written in Scala that runs on top of Apache Spark
- It has numerous Transformers and Estimators that make use of feature abstractions to automate feature engineering, feature validation, and model selection
- Interface: APIs
- Core Algorithms: transmogrifAI model selector trains a few different models simultaneously and use their average validation error to select the final model

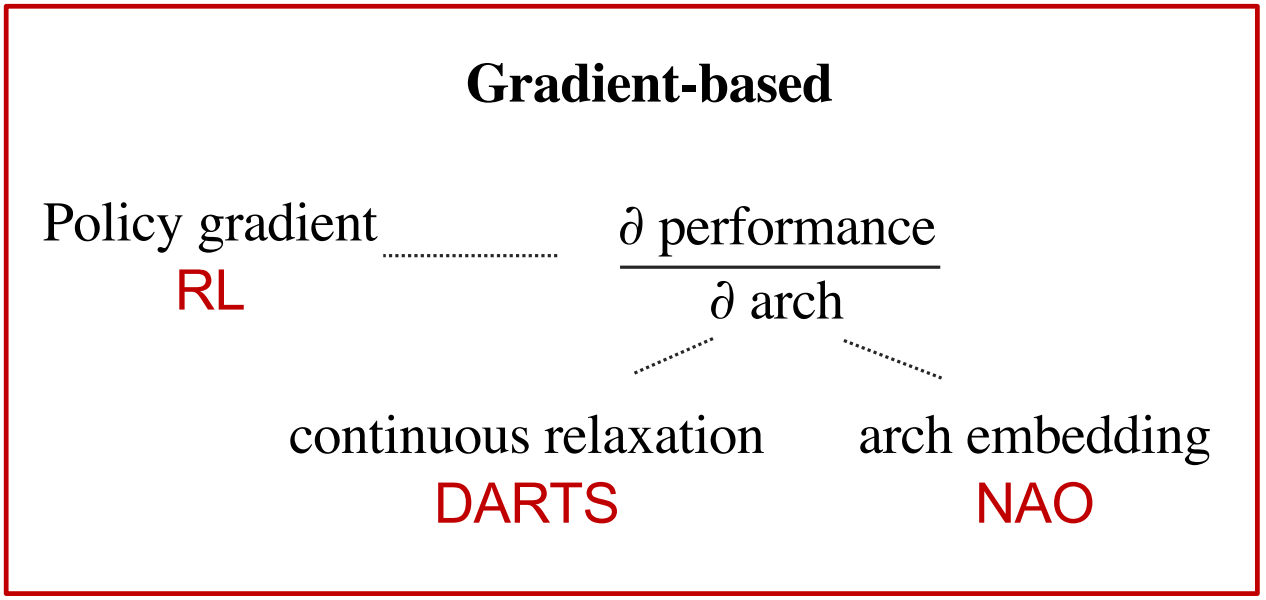
State-of-the-art Methods in Auto Deep Learning

- Neural Architecture Search: Aiming for the highest accuracy on a particular dataset, it searches for deep neural network architectures and automates the optimization of their parameters



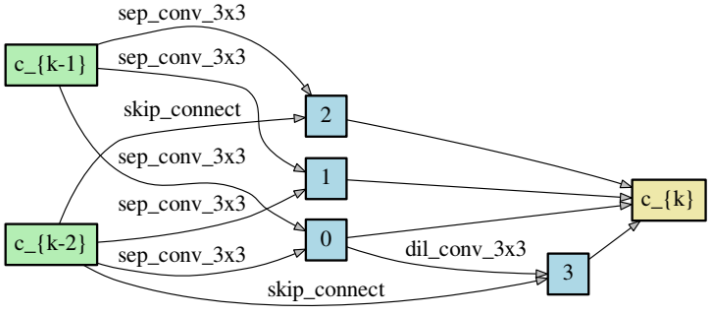
A Neural Unit Designed by NAS

Differentiable NAS

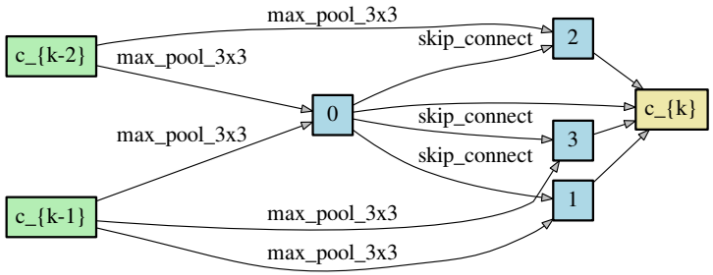


Differentiable NAS

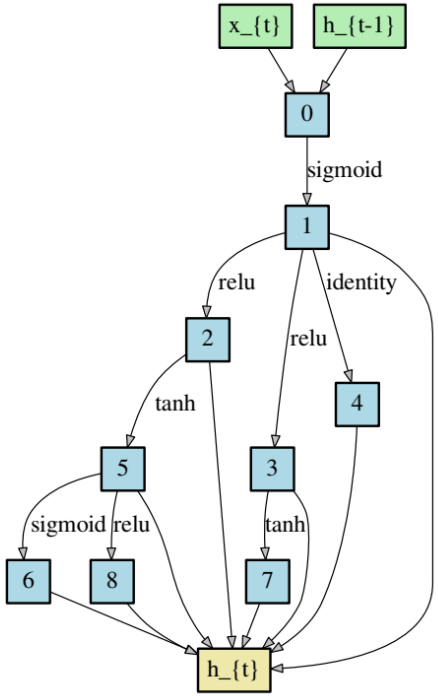
- Architectures Discovered by DARTS (Liu *et al.* ICLR' 19)



Normal Cell



Reduction Cell



Recurrent Cell

Baidu AutoDL

- Use Deep Learning to learn to build DL architectures
- Enables fast, effective and customized Deep Learning model productions
- AutoDL has three components
  - AutoDL Design: Automates the design of a neural network
  - AutoDL Transfer: Supports building networks for small datasets
  - AutoDL Edge: Supports AI + IoT

Milestones

- July 2018: AutoDL 1.0 was released in Baidu Create 2018
- November 2018: AutoDL 2.0 was released in Baidu World Conference
- July 2019: AutoDL 3.0 was released in Baidu Create 2019
- AutoDL has been incorporated Baidu AI platform and multiple products, including:
  - EasyDL, AI Studio, Jarvis, BML

EasyDL
http://ai.baidu.com/ezdl/

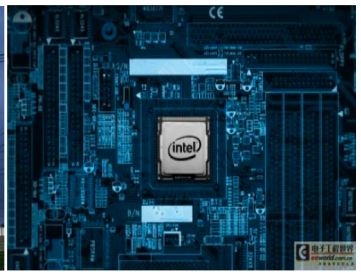
AIStudio
http://aistudio.baidu.com

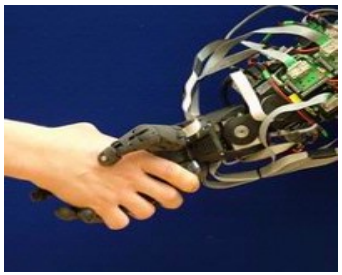
Jarvis
http://di.baidu.com/product/jarvis?castk=LTE%3D#solution-sec0

Summary

- We are in the early stage of the fourth Industrial Revolution; Big Data and Artificial Intelligence are the core power in this era
- In Big Data Lab, we focus on analyzing big data and deep learning theories, in hope of efficiently transforming big data into practical knowledge
- In light of the existing or upcoming trend of using AI, we propose “open and inclusive AI” which could be applied in different areas. Through automated deep learning, we hope to benefit users across disciplines.

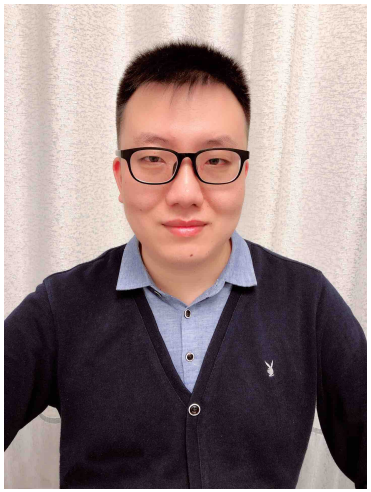


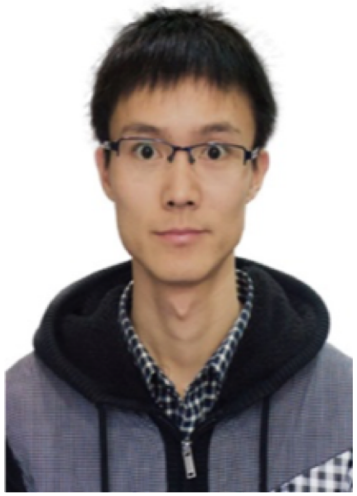




Next

- Neural Architecture Search (NAS) (Siyu Huang)
- Transfer Learning (Xingjian Li and Xuhong Li)
- AutoDLPaddlePaddle: Open Source Platform and Cloud Service (Haoyi Xiong)



Thank you

- For more information:
  - Dr. Dejing Dou, Head of Big Data Lab, Baidu Research
  - Email: doudejing@baidu.com