Building Better Connected World with Artificial Intelligence Technologies

Hang Li
Noah’s Ark Lab
Huawei Technologies
Talk Outline

• Research on AI at Noah’s Ark Lab
• Deep Learning for Natural Language Processing: Advantages and Challenges
Huawei’s Vision:
Building A Better Connected World
THE FUTURE OF ICT WILL DEPEND ON BIG DATA
- MR. ZHENGFEI REN
RESEARCH AREAS

Intelligent Telecommunication Networks

Speech and Language

Recommendation and Search

Big Data Analytics
RESEARCH AREAS

Computer Vision

Internet of Things

Intelligent Devices

Smart City
Intelligent Telecommunication Networks

Software-defined Networks

Network Maintenance
Software Defined Networks

• Goal
  – Build ‘brain’ of network
  – Significantly improve efficiency of SDN using machine learning

• Challenges
  – Network condition dynamically changes
  – Difficult to find optimal policy to control SDN

• Our technologies
  – Flow prediction using online Gaussian process regression
  – Co-flow prediction using clustering
  – Routing using reinforcement learning
  – Prototype system developed
Demo: Network Mind
– SDN Controller Using Machine Learning
Speech and Language

Speech Recognition

Dialogue

Machine Translation
Natural Language Dialogue

• Goal
  – Realizing task-dependent natural language dialogue

• Problems
  – Retrieval based approach
  – Generation based approach
  – Question answering from knowledge base

• Solutions
  – Retrieval based approach using 2D-CNN, state of the art
  – Generation based approach using RNN, first in the field
  – Question answering from knowledge base using deep neural networks
Demo: Generative Approach to Natural Language Dialogue
Neural Responding Machine

太羡慕你了祝旅行愉快

• Using both local and global attention mechanisms
Recommendation
- AppStore Recommendation

• Goal
  – Recommend millions of apps, games
  – To 300 million users

• Challenges
  – Large scale
  – Sparse data

• Our system
  – Online distributed recommendation engine
  – Billions of features and billions of instances
  – State-of-the-art technologies in production
Search
- AppStore Search

• Goal
  – Search millions of apps, games
  – To 300 million users

• Challenges
  – Large scale
  – Ambiguous queries

• Our system
  – Search system using learning to rank technologies
  – State-of-the-art technologies in production
Demo: AppStore Recommendation and Search
Personal Information Management - Image Search

- **Problem**
  - Search pictures on phone or cloud
  - Through speech or text

- **Challenges**
  - Multi modal image retrieval
  - Semantic matching is needed

- **Our system**
  - Deep matching model between text and image, outperforming all existing methods
Demo: Image Retrieval
Multimodal CNN

- One Convolutional Neural Network represents image
- One Convolutional Neural Network represents text
- Multi Layer Perceptron conducts matching

```
a      dog    is  catching  a     ball
```

Diagram: CNN and MLP connections with labels for image and text representation.
Big Data Analytics

Huawei Big Data

Telco Big Data
Telco Big Data

• Goal
  – Enhancing business of telco companies by leveraging big data

• Problems
  – Churn prediction and user retention
  – User location identification using mobile data

• Our solutions
  – Churn prediction using random forest, providing generic tools
  – User location identification using regression model, best performances

BSS (business support system) data
OSS (operation support system) data
Demo: Identifying Location of User with Telco Big Data
Talk Outline

• Research on AI at Noah’s Ark Lab
• Deep Learning for Natural Language Processing: Advantages and Challenges
Fundamental Problems of Statistical Natural Language Processing

• Classification: assigning a label to a string
  \[ S \rightarrow C \]

• Matching: matching two strings
  \[ S, t \rightarrow \mathbb{R}^+ \]

• Translation: transforming one string to another
  \[ S \rightarrow t \]

• Structured prediction: mapping string to structure
  \[ S \rightarrow S' \]

• Stochastic control process: choosing an action in a state, where environment randomly changes states
Fundamental Problems of Statistical Natural Language Processing

• Classification
  – Text classification
  – Sentiment analysis

• Matching
  – Search
  – Question answering
  – Dialogue (single turn)

• Translation
  – Machine translation
  – Speech recognition
  – Hand writing recognition
  – Dialogue (single turn)

• Structured Prediction
  – Named entity extraction
  – Part of speech tagging
  – Sentence parsing
  – Semantic parsing

• Stochastic Control Process
  – Dialogue (multi turn, task dependent)
Advantages & Challenges of Deep Learning

• **Strength**
  – Good at *pattern recognition* problems
  – Data-driven, performance is high in many tasks
  – End-to-end training, little or no domain knowledge is needed in system construction
  – Representation learning, possible in cross modal processing
  – Gradient-based learning, learning algorithm is simple
  – Powerful for supervised learning setting

• **Weakness**
  – Not good at *inference and decision* problems
  – Data-hungry and thus is not suitable when data size is small
  – Difficult to handle long tail phenomena
  – Model is usually a black box and is difficult to understand
  – Computational cost of learning is high
  – Unsupervised learning methods are needed
  – Still lack of theoretical foundation
End-to-End Training
Generation-based Dialogue

Target Sentence

Decoder

Internal Representation

Encoder

Attention Mechanism

Source Sentence
Representation Learning
Symbolic Matching Models

Vector Space Model, BM25, Language Model for IR

Linear Projection into Latent Space
Neural Matching Models

Multimodal Match Model (CNN)

Neural matching models are natural extension of symbolic matching models
Inference and Decision
Comparison between Single-turn QA and Multi-turn QA by Humans

<table>
<thead>
<tr>
<th>Single-turn QA</th>
<th>Multi-turn QA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Q</strong>: How tall is Yao Ming?</td>
<td><strong>Q</strong>: How tall is Yao Ming?</td>
</tr>
<tr>
<td><strong>A</strong>: He is 2.29m tall.</td>
<td><strong>A</strong>: He is 2.29m tall.</td>
</tr>
<tr>
<td><strong>Q</strong>: Who is taller, Yao Ming or Liu Xiang?</td>
<td><strong>Q</strong>: Who is taller, Yao Ming or Liu Xiang?</td>
</tr>
<tr>
<td><strong>A</strong>: He is taller, and Liu Xiang is only 1.89m tall.</td>
<td><strong>A</strong>: He is taller, and Liu Xiang is only 1.89m tall.</td>
</tr>
</tbody>
</table>

- Single turn QA is only related to fact retrieval and answer generation.
- Multi-turn QA needs fact retrieval and answer generation, as well as other processing such as inference and dialogue management. More modules in human brain are involved.
Deep Learning and Multi-turn Dialogue

• DL may not be enough for natural language dialogue
• Key is dialogue management, including dialogue control and dialogue modeling
• Involvement of multiple “modules”, each having multiple “states”
• Recent work tries to use reinforcement learning
• There are many open questions
Long Tail
Data in Natural Language Processing Follows Power Law Distribution

![Graph showing vocabulary size increases when data size increases]

Vocabulary size increases when data size increases

Data from Xiao Chen
Theoretical Analysis
Generalization Ability of Deep Learning

• In practice, usually both training errors and test errors are small, i.e., no over-fitting
• Neural networks can “memorize” training instances
• On the other hand, neural networks can over-fit (i.e., test errors are large although training errors are small), if random noise is injected into training data, cf. Zhang et al. 2017
• Number of parameters is larger than number of training instances
• Many open questions about the learning ability of deep learning

Summary

• Noah’s Ark Lab is conducting research on AI, including
  – Recommendation and search
  – Speech and language
  – Computer vision
  – Big data analytics
  – Intelligent networks
  – Smart city
  – Internet of things
  – Intelligent devices

• Goal is to build better products for Huawei
Summary (cont’)

• Deep Learning is making significant impact to Natural Language Processing

• Advantages
  – End-to-end training
  – Representation learning

• Challenges
  – Inference and decision
  – Long tail
  – Theoretical analysis
Thank you