Introduction to Research at Noah’s Ark Lab
Noah’s Ark Lab

- Research Areas
  - Machine Learning
  - Data Mining
  - Speech and Language Processing
  - Information and Knowledge Management
  - Intelligent Systems
  - Human Computer Interaction

- Founded in 2012
- Located in Hong Kong and Shenzhen
- Collaborations with Universities in Hong Kong
  - A large number of researchers earned PhD at the universities
  - Many interns
  - Strong connections with professors and students
We want to build

Intelligent Mobile Devices

Data Mining & Artificial Intelligence

Intelligent Enterprise

Intelligent Telecommunication Networks
Intelligent Telecommunication Networks

- Software-defined Networks
- Network Maintenance
- Network Planning and Optimization
Stream Data Mining

• **Challenge**
  - Velocity is most salient property of telecommunication data
  - Example: data in datacenter flows in several TBs per second
  - To predict data flow in SDN

• **Solution**
  - Stream Data Platform: StreamSmart
  - Processing one million events in one second on single machine
  - Twice as fast as Storm

• **Technology**
  - Auto load balancing
  - Distributed auto recovery
  - Shared ring buffer
  - Light-weight distributed computation
Graph Data Mining

• **Challenge**
  - Graph data is everywhere
  - Scale is large
  - Predict data flow in of network

• **Solution**
  - Graph data mining platform: VENUS
  - PageRank on Twitter graph, almost one times faster than GraphChi on single machine
  - Vertex-centric streamlined processing
Intelligent Enterprise

- Supply Chain Management
- Customer Relationship Management
- Human Resources Management
- Information and Knowledge Management
- Communication
Telco Big Data

**Challenge**
- Volume is extremely large
- Consisting of BSS data and OSS data

**Solution**
- Example: churn prediction
- Retaining customers is key for telco companies
- Prediction rate = 90% for top 10K users
- Increased retention rate from 15% to 35%
- Extracting useful features from both OSS and BSS data
Intelligent Banking

**Case study:**
- Commercial Bank A
- Traditionally, less than one hundred features
- Currently, thousands even millions of features developed by Huawei’s tool
- Increased by a factor of six the success rate of targeted advertisements.

**Challenge**
- Data sparse, scale is large

**Solution**
- Automatic feature engineering
Intelligent Mobile Devices

- Information Recommendation
- Information Extraction
- Personal Information Management
- Machine Translation
- Natural Language Dialogue (Question Answering)
Intelligent Help

• **Feature**
  - Asking questions about usages of Huawei phone in natural language

• **Challenge**
  - Mismatch between question and answer
  - “How can I access the internet?” vs “Ways of connecting to the web”

• **Solution**
  - Learning to rank and learning to match technologies
  - Recognize more than 90% of the questions and accuracy of answers is over 92%
  - Currently 100K questions / day
Natural Language Dialogue

• Challenge
  • Natural language understanding is AI complete problem
  • Turing Test

• Opportunity:
  – Huge amount of conversation data is available
  – Computers becoming more powerful
  – Deep learning technologies available

• Solution
  • Retrieval based approach and generation based approach
  • Retrieval: deep matching models
  • Generation: recurrent neural network
  • Accuracy: 70%+
  • Organizing Short Text Conversation (STC) Competition
Take-away Messages

• Noah’s Ark Lab is working on data mining and artificial intelligence
• Doing research to build
  • Intelligent telecommunication networks
  • Intelligent enterprise
  • Intelligent mobile devices
• Has already made many significant achievements
• We look forward to having more collaborations with the local research communities
• We are hiring employees and interns
For more information, please visit http://www.noahlab.com.hk/
Any questions, please contact us at noahlab@huawei.com