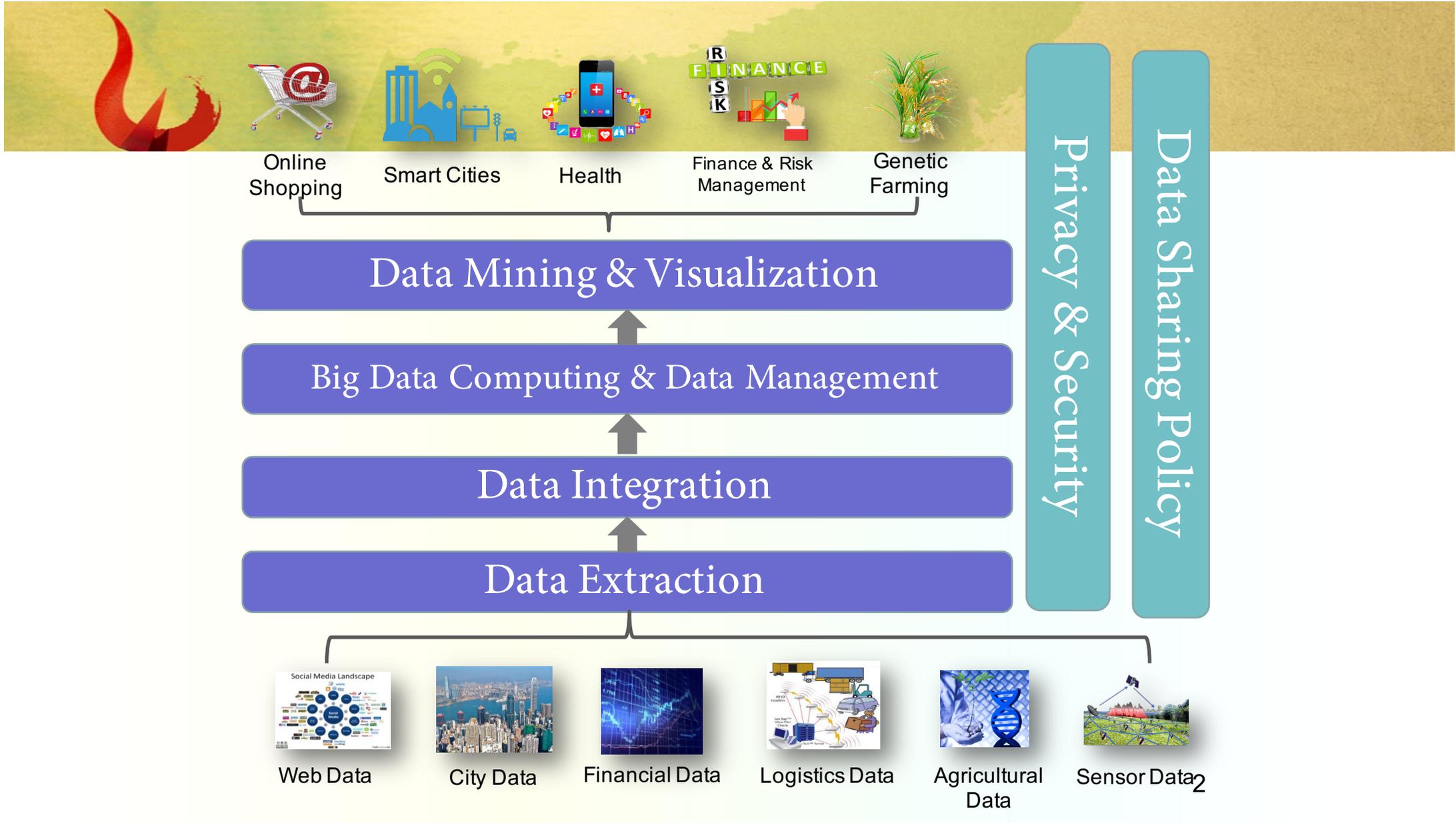


HKUST Big Data Institute (BDI)

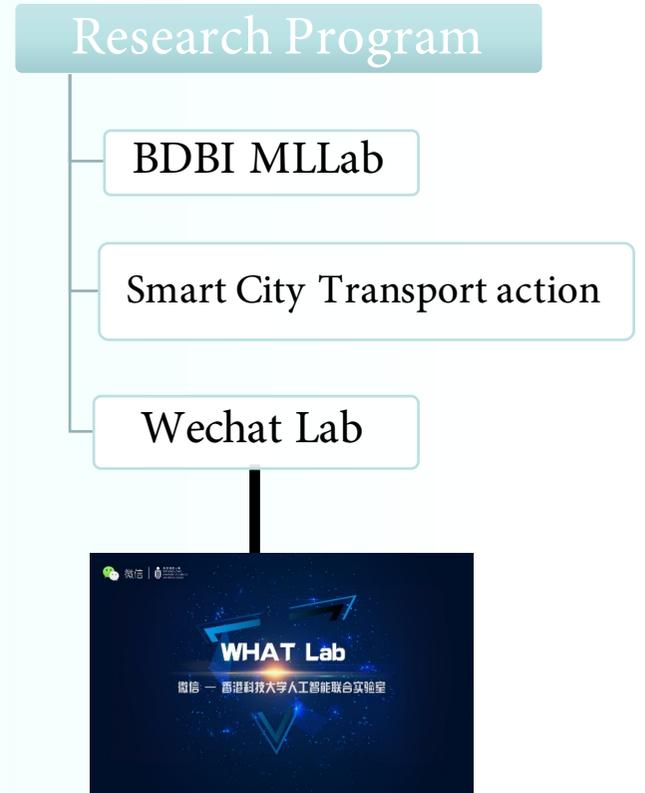
Lei Chen and Yang Wang
Hong Kong University of Science & Technology





BDI Research Program Structure

- **WeChat-HKUST Joint Lab on Artificial Intelligence (WHAT Lab)**
 - aims at social networking big data mining and machine learning, natural language processing and robotics research.
- **Big-Data Bio-Intelligence (BDBI) and Machine Learning Lab**
 - aimed at the development of advanced machine learning systems and promoting applications of machine learning in bio and genetic areas.
- **Smart City Transportation Lab**
 - aimed at using AI and Big Data technologies in smart city applications.





World Experts in Big Data and Artificial Intelligence Gather at HKUST to Share Insights into Future



Dr Tieniu Tan, Vice Minister of the Liaison Office of the Central People's Government in HKSAR (7th from right), HKUST President Prof Tony F Chan (6th from right), Vice-President for Research and Graduate Studies Prof Nancy Ip (7th from left), Dean of Engineering Prof Tim Cheng (5th from right), Prof Qiang Yang, Head of the Department of Computer Science & Engineering and Director of the Big Data Institute (4th from right), with world-class academic and industry leaders in Big Data and AI fields, and HKUST faculty [\[Download Photo\]](#)



Joint Research Projects

- WHAT Lab: Tencent entered into a 5-year research partnership with BDI since late 2015, amounts to RMB10 million in total; 11 faculty; 20 students
- BDBI-Machine Learning Lab: Donation by Mr. Raymond Chu for a duration of 5 years, totals US\$1 million; 3 faculty, 1 RaP, 16 students
- Smart City: ITF(Innovation and Technology Fund) project partnered with Digital China, HK\$20 million for 2 years; 12 faculty; 14 students
- Smart Transportation: ITF project partnered with Thales, HK\$10 million for 2 years; 10 faculty; 56 students

面向高维度的机器学习的计算框架-Angel

Original 2016-05-01 腾讯大数据 腾讯大数据

简介

为支持超大维度机器学习模型运算，腾讯数据平台部与香港科技大学合作开发了面向机器学习的分布式计算框架——Angel 1.0。

Angel是使用Java语言开发的专有机器学习计算系统，用户可以像用Spark, MapReduce一样，用它来完成机器学习的模型训练。Angel已经支持了SGD、ADMM优化算法，同时我们也提供了一些常用的机器学习模型；但是如果用户有自定义需求，也可以在我们提供的最优化算法上层比较容易地封装模型。

Angel应用香港科技大学的Chukonu 作为网络解决方案，在高维度机器学习的参数更新过程中，有针对性地给滞后的计算任务的参数传递提速，整体上缩短机器学习算法的运算时间。这一创新采用了香港科技大学陈凯教授及其研究小组开发的感知上层应用（Application-aware）的网络优化方案，以及杨强教授领导的大规模机器学习研究方案。



WHAT Lab: A. Establishment

WHAT LAB, which is short for WeChat-HKUST Joint Lab on Artificial Intelligence Technology, is dedicated to foster artificial intelligence and big data research to improve people's living and advance the frontiers of knowledge, marking a milestone in the collaboration of WeChat and the higher education sector.

WeChat and HKUST will jointly conduct Artificial Intelligence (AI) Technology related research and explore the far-reaching frontiers of AI. This collaboration on AI research is expected to be long-term and world-leading. Research areas of WHAT LAB include intelligent robotic systems, natural language processing, data mining, speech recognition and understanding.



2015-11-26 WHAT Lab was established in HKUST



WHAT Lab: B. Research Area

1. Natural language processing
2. Data Mining & Visualization
3. Video Analysis
4. Large -Scale Machine Learning
5. Robotic Application

Scale :

1. 11 projects
2. 11 professors from CSE, ECE, MATH, ISOM
3. 20 students





WHAT Lab: C. Achievement I- Machine Reading



- Machine Reading aims to develop Machine Learning algorithms that could read and comprehend natural language documents as humans do. With Machine Reading, natural language information is converted to the form that could be processed by computers, and could be further utilized in applications such as summarization, question answering and dialogue system.



Xtecher

最初在腾讯内部提出要围棋AI的姚星，一开始就表示，围棋AI只是一部分，绝艺背后“精准决策”AI能力完全可以用在无人驾驶、量化金融、辅助医疗等领域。从围棋入手AI研究，在不断深入后，腾讯在其他领域AI应用必然将被逐渐挖掘。



“绝艺”横空出世，腾讯野心几何？

市场和人群性别偏见人们可能会认为股票市场是一种“群体智慧”的产物，以为公司估值是基于所有投资者的集体看法。但通过大量计算的研究后，人们发现即使个人没有偏见，如果他们相信其他人有偏见，这就会让市场价格产生偏差。



机器如何帮助审查“群体智慧”的偏见？ | Xtecher硅谷

那么，谈及DeepMind的医疗区块链时，不同参与者之间的权益分配又是怎样的呢？基础数据属于医院，而软件则属于DeepMind，Suleyman说。DeepMind表示，软件处理患者数据时创建的数据日志也由医院管理。



机器人先律己：谷歌的DeepMind开发自身盗取用户医疗数据技术 | Xtecher硅谷

我们仍处于AI前夜。“人民日益增长的人工智



WHAT Lab: C. Achievement II- dialog system

- A dialog system is a computer system intended to converse with a human. Through the dialogue system, computer information can be translated into natural language description, and human language can also be translated into computer information, so as to achieve human-computer interaction.
- Reinforcement and transfer learning, can be used to solve the problem of dialogue system. Reinforcement learning can solve the problem of delayed feedback in multiple rounds of dialogue, while transfer learning can help target areas by using data from similar fields, which can solve the problems of traditional intensive learning data.





WHAT Lab: C. Achievement IV-Model-based Global Localization for Aerial Robots using Edge Alignments

Model-based Global Localization for Aerial Robots using Edge Alignment

Kejie Qiu, Tianbo Liu and Shaojie Shen



High resolution video available at:

<http://www.ece.ust.hk/~eeshaojie/ral2017kejie.mp4>

- The video contains three parts
- The first part presents the localization accuracy and global consistency by comparing with the ground truth provided in the indoor environment.
- The second part shows the real-time localization results in outdoor case.
- The third part of the video shows the closed-loop control by a trajectory tracking experiment using the proposed method for 10 state feedback."



BDBI-Machine Learning Lab: A. Establishment

The Big Data for Bio Intelligence Laboratory (BDBI) aims to become a leading laboratory in the research of big data for biological intelligence and to bridge the knowledge gap between academics and practitioners.





BDBI-Machine Learning Lab: B. Research Areas

- Research areas of the Laboratory include new big data solutions such as “deep learning solution”, which generates rich features to describe a machine learning problem in order to let computers make decisions, and “transductive transfer learning” – allowing computer models to be easily adapted for use in many different application domains. It will also focus on genetic farming with objectives of making the process more automatic and user-friendly and scaling it to cater to very large data sources.





Smart City: A. Needs & Goals



- In this project, we proposed a people-aware smart city framework that integrates data extracted continuously from the people, discovers their needs from integrated multi-source data, and finally determines the best resource allocation plans to satisfy these needs.
- People's needs from the areas of education, health, travel, safety, finance and entertainment, which all have measurable objectives, will be studied in this project.
- In order to achieve the goals in the framework, several state of art techniques will be developed including data integration solutions to handle different data sources with different formats, transfer learning-based mechanisms to reveal knowledge, and machine-human collaborative approaches to 13 make wise decision.



Smart City: B. Project Overview



31°

57°

UV Index: 3
UV radiation: moderate
Wind direction: West
Wind speed: 31 km/h
Wind max gust: 39 km/h
Air pressure: 998.5 hPa
Visibility: 8 km
Rain: 0 mm

A small photograph showing a cityscape, likely Hong Kong, with buildings and water.

- Urban Geographic Information Integration



Smart City: C. Application Area

SMARTCITY BY SMARTCITY GROUP

MENU 

Weather

PROVIDES WEATHER INFORMATION OF HONG KONG

Finance

CONTAINS STOCK INFORMATION OF COMPANIES LISTED ON THE HONG KONG STOCK EXCHANGE

Public Information

INCLUDES TRANSPORTATION, PUBLIC FACILITY, ETC.

Citizen

INCLUDES SOCIAL MEDIA, NEWS, HOUSING, JOB, ETC.



Smart City: D. Open Data

Achievements: Open Data APIs (aggregated with Digital China platform)

 事件搜索 0 0 免费	 风报数据 0 0 免费	 香港求职信息 0 0 免费	 香港房地产 (租房及买房) 信息 0 0 免费
 香港政府公众通知 0 0 免费	 香港公共信息 0 0 免费	 香港医疗 0 0 免费	 香港交通数据 0 0 免费



Smart Transportation: A. Features

Hong Kong University of Sciences and Technology (HKUST) and Thales are carrying out research and development to build a Big Data platform to address two critical problems in the public transport.

The project is unique to the extent that academia, industry and government work closely together to build an interdisciplinary and cross-domain solution for problems pertinent to Hong Kong and other large cities. The Big Data platform helps public transport agencies to build smart transport solutions in moving people efficiently and safely and, hence, enhancing citizen's quality of living in a smart city.



Smart Transportation: B. Purpose and Scope of research

- The purpose of this platform :
 - (1) The platform aims at effectively monitoring and directing the crowd in railway stations so that early warnings can be given on potential dangers.
 - (2) Ensuring smooth operation of railway transport system by predicting potential major equipment failure.
- The platform will cover a number of frontiers of big data research, including data integration, data analysis, human factors, optimization / visualization, transfer of learning, simulation and operational research.





Other participation in cooperation

- Personalized Real-time Air quality Informatics System for Exposure(PRAISE-HK) : a 5-year project funded by the HSBC 150th Anniversary Charity Program (more than HK\$30M)
BDI faculty involved as Co-I: QU Huamin and CHEN Lei
- Learning and Assessment for Digital Citizenship : receives HK\$20M from TRS, 2016-2017
BDI faculty involved as Co-I: TC Pong, Huamin Qu, and Xiaojuan Ma
- Big Data for Smart and Personalized Air Pollution Monitoring and Health Management: receives HK\$50M, 2017-2018
BDI faculty involved as Co-I: Huamin QU
- Evidence Based Education based on Data Analytics-Moving Towards a 360 View of the Students.
BDI project involved: Lei Chen and Qiang Yang
- Data mining and deep learning of the human face for enhanced social power and advertising effectiveness: This project is funded HK\$900K under University-Industry Collaboration Programme (UICP).
BDI faculty involved: Huamin Qu (PI)



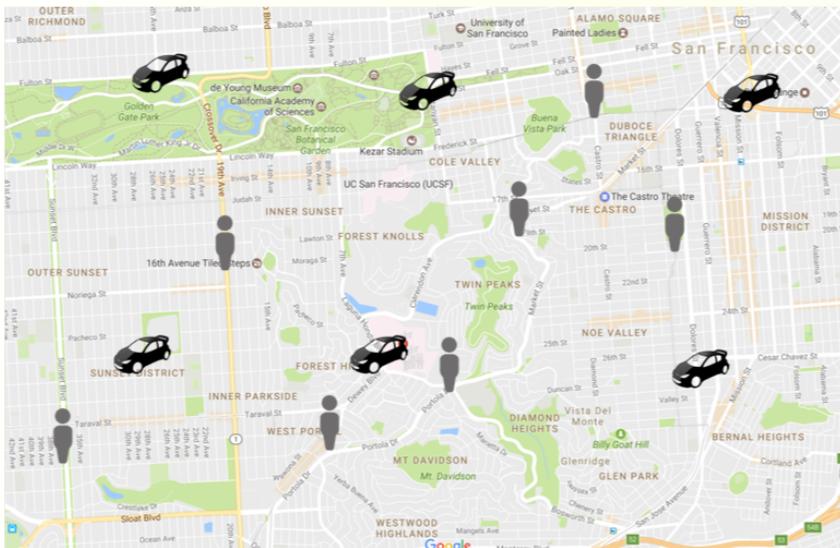
Research Achievements

- Prof Ke Yi received ACM SIGMOD Best Paper Award 2016 for [Wander Join: Online Aggregation via Random Walks](#), and ACM SIGMOD Best Demonstration Award 2015 for STORM: [Spatio-Temporal Online Reasoning and Management of Large Spatio-Temporal Data](#).
- BDI Acting Director Professor Lei Chen is the Editor in Chief of the VLDB journal and Associate Editor in Chief of the IEEE TKDE journal.
- BDI and WHAT Lab member Professor Kai Chen and Tencent launched a new big-data and AI platform known as Angel and Amber to tackle the fast-growing big data analytic demands at WeChat and Tencent Cloud.
- BDI and WHAT Lab member Professor Huamin Qu's WeSeer system is being used to analyze the information-diffusion patterns of millions of news articles each day.
- Best Presentation Award — Prof Dekai WU and Kartek ADDANKI. [Neural Versus Symbolic Rap Battle Bots](#). 41st International Computer Music Conference (ICMC 2015), Texas: Sep 2015
- Prof Jiguang Wang and others' paper [Spatiotemporal genomic architecture informs precision oncology in glioblastoma](#) is published in Nature Genetics in March 2017
- Prof Yangqiu Song others' paper [HinDroid: An Intelligent Android Malware Detection System Based on Structured Heterogeneous Information Network](#) won the Best paper and best student paper in applied data science track, in ACM SIGKDD 2017.

Research Achievements

- **Smart Ridesharing:**

With the cooperation of DiDi Chuxing (the biggest online car-hailing company in Mainland China), we can utilize the huge amount of data generated by millions of drivers and customers to help the company improve the efficiency of their services and the user experience of both drivers and customers. We help DiDi to design smart vehicle dispatching strategies and dynamic pricing strategies such that the efficiency of the ridesharing service can be improved, and the overall profit of the platform can increase.





Research Achievements

- **WHAT Lab: WeChat Crowdsourcing Platform**

The WeChat Crowdsourcing Platform is designed and developed by WHAT Lab and WeChat team together. Researchers can publish crowdsourcing tasks on the platform and WeChat users can participate in and get monetary rewards. Different mechanisms for task assignment and answer aggregation are equipped and plenty of real tasks from HKUST and Tencent have been published and finished on the platform. It is both a useful tool for data labeling and an industry level crowdsourcing research environment.



Scan with WeChat

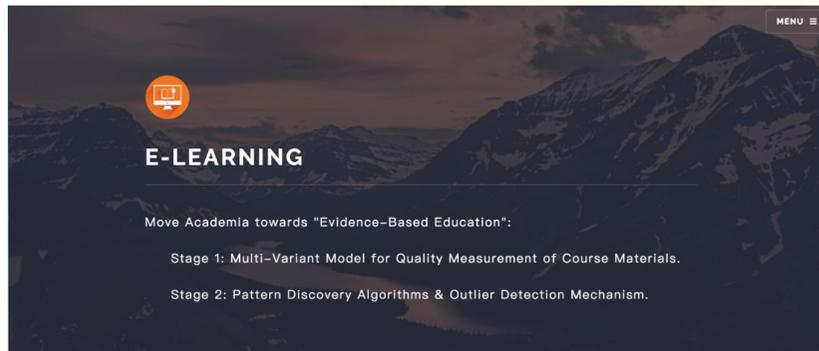


Research Achievements

Mixture Multi-Variant Quality Measurement Model;

Pattern Discovery Algorithms;

Effective Outlier Detection Mechanisms



Research Achievements

- **WHAT Lab: WeSeer System: (go online)**

The WeSeer system developed by WHAT Lab was deployed and applied to WeChat, Tencent for daily propagation analysis. The system enable to analysis how official public account article information propagate in WeChat platform from different perspectives, involving a 3D global overview, time-varying propagation view, community detection view, etc.

微信小秘密: 2016 年那些 10w+ 文章是怎么刷爆朋友圈的?

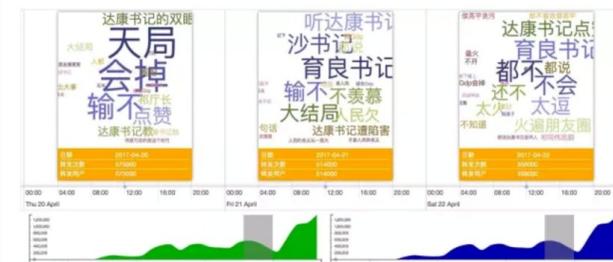
Original 2016-12-30 WeChat TechPower WeChat TechPower



这次微信人工智能实验室 WHAT Lab 用数据可视化算法也看了一回《人民的名义》

大结局的前一周
《人民的名义》相关热词曲线

4月20日~4月22日



2017年4月20日, 带有《人民的名义》标题的微信文章转发次数达到了575000次, 朋友圈几乎被以达康书记为首的汉东男子天团所占领, 可以看到很多人对大结局充满了期待。



Industry & Academic partners

□ ITF Project Sponsors:

- ❖ Thales
- ❖ Digital China

□ Tencent WeChat

- ❖ Established a joint lab (WHAT Lab) with HKUST in December 2015
- ❖ WHAT Lab Opening Workshop in 26 November 2015
- ❖ WHAT Lab one year anniversary in 16 November 2016

□ Shanghai Jiaotong University

- ❖ Joint workshop on big data in September 2016

□ NVIDIA Deep Learning Institute

- ❖ Joint workshop on practical deep learning in July 2017

□ Microsoft Research Lab Asia

- ❖ HKUST-MSRA Cup on “Big Ideas of Artificial Intelligence” in December 2016

BDI Seminars



In collaboration with different departments and divisions, HKUST BDI has been organizing a monthly lecture series from January 2017 onwards, inviting cross disciplinary speakers to enhance students, teaching staff and public's understanding on related big data topics and foster exchange and collaboration between academia and industry beyond borders and research fields.



BDI Workshop



HKUST BDI is excited to partner with NVIDIA Deep Learning Institute, which has established itself as the leading provider of processing power for AI software, in launching its first-ever practical deep learning workshop on 27 July 2017.



BDI Management

- **BDI Advisory Committee (Internal):**
 - Chairman: Dean of School of Engineering: Professor Tim Cheng
 - Dean of School of Science: Professor Yang Wang
 - Dean of School of Business & Management: Professor Kar Yan Tam
 - Associate Vice President of Knowledge Transfer: Professor Enboa Wu
 - Head of Department of Industrial Engineering & Logistics Management: Professor Guillermo Gallego
 - Head of Department of Electronic and Computer Engineering: Professor Bert Shi
 - Head of Division of Biomedical Engineering: Professor I-Ming Hsing
 - Head of Division of Social Science: Professor Kellee Tsai
 - Acting Director of HKUST Big Data Institute: Professor Lei Chen
 - Associate Director of HKUST Big Data Institute: Professor Yang Wang
- **BDI Advisory Committee (External):**
 - Mr. Francis Kwok, Founder & Chief Product Designer of Radica Systems Ltd
 - Mr. Herbert Chia, Venture Partner of Sequoia Capital China
 - Professor Michael Franklin, Liew Family Chair of Computer Science, Senior Advisor to the Provost for Computation and Data, Interim Director of Computation Institute, University of Chicago
 - Professor Leonidas J. Guibas, Professor of Computer Science, Stanford University
 - Professor Tamer Ozsu, Professor of Cheriton School of Computer Science, University of Waterloo



BDI Management

The Executive Committee

- **Members:**
 - Professor Lei Chen, Acting Director of HKUST Big Data Institute, and Department of Computer Science and Engineering
 - Professor Yang Wang, Associate Director of HKUST Big Data Institute, and Dean of Science
 - Professor Cameron Campbell, Professor of Division of Social Science
 - Professor Inchi Hu, Chair Professor of Department of Information Systems, Business Statistics and Operations Management



Publicity & outreach

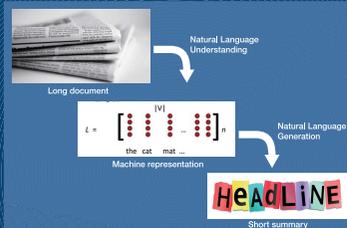
- Website development and maintenance (ongoing)
- Demos for constant visitors
- Big Data & AI Day (26 May 2017) – open conference with nation-wide participation among academia and industry
- Monthly open seminars series – partners with different departments on cross-disciplinary topics
- Organize student competitions (e.g. MSRA Cup, Smart City East Kowloon Competition, etc.)
- Researchers in projects take part in external competition to promote research strengths (e.g. entry for Hong Kong ICT Awards 2017 - Best Smart Hong Kong Award, 2017 Shenzhen City Open Data Innovation, etc.)

Abstractive Summarization using Sequence-to-sequence RNNs and Reinforcement Learning

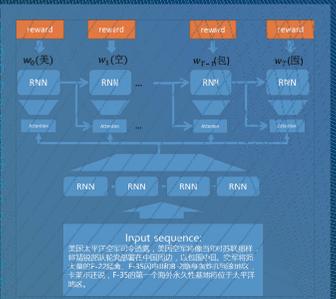
By Yuxiang WU

Automatic summarization aims to produce a short summary for a given long document, such as news article, academic paper or even source code. When applied in social media, it may be able to produce short but informative gist of the news posts so that readers could save some precious reading time.

Recently, some large-scale abstractive summarization datasets are released, and therefore it becomes feasible to train an abstractive summarization model. We built a summarization system based on sequence-to-sequence model, and we exploit Reinforcement Learning to enhance the performance. The experiment result shows that sequence-to-sequence model trained with Reinforcement Learning achieves performance (evaluated by ROUGE) similar to current state-of-the-art CopyNet, which has a complex copy mechanism.



With the surge of social network and mobile Apps, smartphone users are usually overwhelmed by news posts. Users could not read all the news articles in detail and may not even be able to open them. If we could automatically summarize the long articles into short summaries, we could shorten the reading time taken for each post while keeping the users informed with key information. Before producing an informative gist, our algorithm must understand what the article says. This is called Natural Language Understanding, which transforms text into a machine representation. Then, it tries to identify the key information through this representation, and generate natural language output. Recent advances in Deep Learning improve machine's capability in natural language understanding and generation. Furthermore, some large-scale summarization datasets are published in the past two years, and therefore it becomes feasible to train a deep learning model for abstractive summarization.



原文：美剧《生活大爆炸》中，“谢耳朵”威尔·肖伊特勒的化身，《穿越》中男主角艾克曼和艾莉莎的性：女主角最喜欢的小机器人果巴变成了萌萌的完全体。TVC新剧《穿越》也吸引了机器人。设计显示，美国影视剧给中国品牌带来了直接利益。《国际商报》

原标题：中国品牌“融入”美国影视剧 增强学习：美国影视剧给中国品牌带来直接利益

原文：7日上午，国内14个城市的不同地区上，5万多名广场舞大妈，发起广场舞“占领”行动。挑战最大规模广场舞(多场)吉尼斯世界纪录成功。创下新的世界纪录，也是全国广场舞爱好者的“狂欢盛会”。

原标题：厉害了我们的广场舞 增强学习：国内14个城市，5万广场舞大妈挑战最大规模广场舞

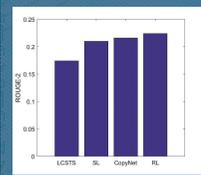
Our summarization model is based on Sequence-to-sequence model with attention mechanism. Its input is a news article, namely a sequence of words. The output is the corresponding summary, which is also a sequence of words. It has two parts, an encoder and a decoder. The encoder transforms the input article into a matrix of numbers, representing the semantic information given in the article. The decoder would then produce the summary word by word, based on its semantic representation of the article. Both the encoder and decoder are composed of Recurrent Neural Networks (RNN), and more specifically, Gated Recurrent Units (GRU) or Long Short-term Memory (LSTM). Moreover, attention mechanism is used to locate important information in the article. When decoder generates the next word, it would assign weights to all parts of the article by computing the similarity between decoder's hidden state and the semantic representation. The weighted representation would be fed into the decoder, and therefore it would be more likely to generate sentences related to the key points. We further improve the performance of our model using Reinforcement Learning. Supervised Learning (SL) tries to maximize the likelihood of generating ground truth outputs. Different from SL, the model is optimized to achieve higher generation reward. The reward is defined as the similarity between the generated summary and the real gist.

原文：阿里与优酷土豆宣布建立战略合作伙伴关系。阿里巴巴与优酷基金会12.2亿美元入股优酷土豆。优酷土豆公告显示：阿里巴巴斥资5.5亿美元，12.2%。优酷土豆估值达2.6亿。阿里巴巴持股比例为16.5%。云锋基金持股比例22%。

原标题：优酷土豆被阿里入股 增强学习：阿里与优酷土豆宣布建立战略合作伙伴关系

原文：当地时间14月2日，在巴基斯坦肆虐的印度厄方地区口，发生了一场自杀式炸弹袭击。这起袭击事件造成了至少55人死亡，其中包括了孩子和孕妇。共有1200人受伤。巴基斯坦官员表示，此次的事件是一起自杀式袭击。

原标题：巴基斯坦发生自杀式炸弹袭击 增强学习：巴基斯坦发生自杀式炸弹袭击事件造成至少55人死亡



Model	ROUGE-2 (F1)
LSTMs	0.1742
CopyNet (Char)	0.216
CopyNet (Word)	0.223
Our (SL)	0.2098
Our (RL)	0.2340

Experiments and case study

The diagram and table on the left show the performance of our model compared to the baselines. The performance is evaluated using the ROUGE-2 score, a metric that is commonly used in summarization tasks. Our model with Reinforcement Learning achieves the performance that is competitive with current state-of-the-art model CopyNet. The boxes above plus four pairs of article and summary generated by our RL model. By comparing the summaries with headlines written by human, we could find that they are close in meaning. The generated gist is usually fluent and informative, and has a good coverage of the key points in the article.

References

[1] Liu, C. Y. (2004). Rouge: A package for automatic evaluation of summaries. In Text Summarization: Approaches to Processing of the ACL-04 workshop (Vol. 6).

[2] Gu, J., Lu, Z., Li, H., & Li, W. Q. K. (2016). Improving CopyNet Mechanism in Sequence-to-Sequence Learning. arXiv:1603.06393 [CS]. Retrieved from <http://arxiv.org/abs/1603.06393>

[3] He, B., Chen, Q., & Zhu, F. (2015). LSTMs: A Large Scale Chinese Short Text Summarization Dataset. arXiv:1506.03663 [CS]. Retrieved from <http://arxiv.org/abs/1506.03663>

[4] Rush, A. M., Chopra, S., & Weston, J. (2015). A neural attention model for abstractive sentence summarization. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Retrieved from <http://arxiv.org/abs/1509.00859>

Hong Kong Real Estate Market Analysis

By CHEN, Jiahang, Smart City Lab

Real estate market analysis focus on the values of properties. It helps customer to find the current values of their own properties or the estate with potential value increase or falling. Although individual seldom purchase property, we could have an image of real estate market in a certain time period by combining all purchase behaviors together. This helps the government to know the hottest region of property exchange and have an idea about people's willing on selling or buying real estates.

Historical transaction data reveals the market trend on property value. However, in real world, the value of property is strongly influenced by economic situation, which means the real estate market might have tremendous change when the stock market suddenly fell. To understand the real estate market, we take not only transaction data but also stocks from properties and construction category into consideration.

This project consists of two parts, transaction data visualization and property value analysis based on historical transaction data and stock market situation. We hope this model could give a clear image on Hong Kong property market.



Figure 1, Hong Kong Recent Transaction Data

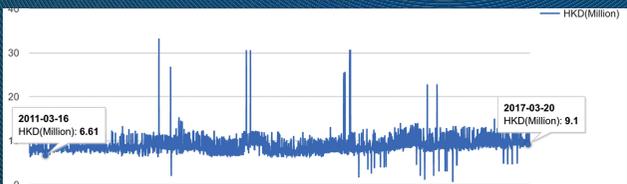
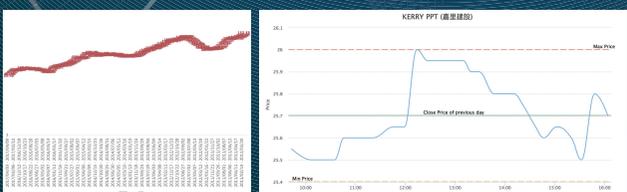


Figure 2, Transaction Data For Festival City (2011-2017)



First part is transaction data visualization. We collect data from big online Hong Kong property portal. After data cleaning and linkage, for each property, we have the gross floor area, net floor area, price, location and transaction date. Figure 1 shows each region of Hong Kong filled with different color based on recent transaction occurred in that place. In this way, user could have a clear view of current property market, for example, which is hottest property-exchange region.

In second part, we take care of historical transaction data and stock market situation. Figure 2 shows the transaction data of Festival City from 2011 to 2017. Even though there was several fluctuations during this period, the whole trend is increase and makes Festival City a good real estate for investment.

Figure 3 is the Genta-City Leading Index, which shows the real estate market situation using data from top-100 real estate transaction data. This index similarly increases during 2011 to 2017.

When implementing stock change into our model, we find out stock price change is hard to follow and make usage of, but we could treat stocks from property and construction category as a whole and find some intuitions.

During the analysis, we find that, besides stock data, newspaper also contains information on real estate market and it works more like a summary or an answer to why the property price changes giving certain circumstances. For big companies or groups, they always move ahead of news release while normal people move afterwards. In further study, we might implement natural language processing algorithms to take newspaper content into consideration and predict value changes once similar news are released.



Future Plans

- AI-FINTECH Program with Business School and Hong Kong Financial Industry
- Student Internship Program with Industry
- Machine Learning Lab: Joint with BDBI Lab
- Joint University Program with Other Top Universities in the World



*APPENDIX:
More Media Coverage
About BDI*



Publicity & media coverage

Wen Wei Po, 6 May 2016

首頁 > 內地 > 正文

港科大聯手微信建「智能實驗室」

時間：2016-07-28 03:16:41 來源：大公網

Ta Kung Pao, 28 July 2016

下一篇：港科大教授打造AI金融「先知」

【大公報訊】記者周琳北京報道：去年底，微信團隊與香港科技大學聯合成立「微信—香港科技大學人工智能聯合實驗室」，由楊強領銜。他對大公報表示，香港在AI領域有很多做機器學習、深度學習、機器翻譯的優秀人才，在此之前他們一直深耕於學術界，希望由此開始打開兩地在人工智能領域的合作。

楊強介紹，香港科技大學在人工智能，機器人和大數據領域在世界上都是處於領先地位。機器學習研究團隊和學生曾經獲得多項世界比賽冠軍，包括ACM KDDCUP大賽、諾基亞大賽以及ImageNet。在遷移學習、統計機器翻譯和深度語言學習領域也引領世界研究潮流。學校研究的情感機器人被「科學美國人」報道，大學出來的創新公司包括世界聞名的「大疆」公司。大學雲集國際著名協會的院士，包括幾十名IEEE Fellow。楊強本人也是國際高等人工智能協會（AAAI）的首名華人院士，2015年國際人工智能大會的主席，以及IEEE大數據期刊的創始主編。

目前，推動微信新技術和功能研發的核心是成立於2010年9月的「模式識別」團隊，研究方向包含語音識別、圖像識別、語義理解等，相當多的基礎研究成果（如語音轉文字）已被集成為微信功能。而與香港科技大學聯合建立人工智能實驗室，無疑將極大提升微信在人工智能領域的技術實力。

聯合實驗室研究範疇包括機器學習、自然語言處理、機器人、數據挖掘及語音辨識，將使用微信平台上6.5億月活躍用戶和千萬級公眾號的數據資料。楊強說，微信將在實驗室投入一千萬支持研發，但最為難得的是微信開放了所有的數據，不涉隱私地做研究。

科大設大數據生物智能實驗室



■左起：陳繁昌、朱慧恒、副校長(研發及研究生教育)李行偉、楊強合照留影。科大供圖

香港文匯報訊（記者高鈺）香港科技大學昨正式成立大數據生物智能實驗室，為生物學及醫療保健方面設計大數據分析方案，實驗室將由科大新明工程學教授、計算機科學及工程學系講座教授兼系主任楊強，及數學系講座教授兼系主任汪揚共同領導。

科大校長陳繁昌特別鳴謝實業家朱慧恒的捐助，又指大數據的應用將會為人類的生活帶來革命性的轉變，有關研究是該校策略發展方向之一，而隨着實驗室的成立，他深信科大能為大數據研究的迅速發展帶來新見解。

大數據生物智能實驗室的研究範疇包括「深度學習方案」，即透過豐富功能描述機器的學習問題，從而讓電腦作出決定；以及能讓電腦模型輕易地在不同領域中應用的「直推式遷移學習」。實驗室亦會專注研究基因養殖，令過程更為自動化和易於使用，以及配合龐大的數據來源。

人民網 香港

香港致力打造大数据驱动的智慧城市

2016年06月28日12:42 来源：新华社

分享到

原标题：香港致力打造大数据驱动的智慧城市

“我们的口号是要把香港打造成大数据驱动的智慧城市。”香港科技大学计算机科学与工程学系主任杨强对新华社记者说，在他的牵头下，科大组织了一批专家开展智慧城市建设和大数据的应用。

香港特区政府行政长官梁振英在《2016年施政报告》中也提到，特区政府将与科研及私营机构共同研究建设智慧城市。特区政府创新和科技局（创科局）将负责制定智慧城市的数码架构和标准。

创科局回应记者关于香港智慧城市构想时表示，将于2016年下半年开展一项顾问研究，为香港制定一个整体框架作为智慧城市发展蓝图，研究范围包括策略、发展计划、管治安排、数码架构和标准、公私营合作模式等。目标是香港制定一个至2030年的智慧城市长远发展计划，利用创新及科技解决面对的都市挑战，并提升城市管理和改善市民生活质量。顾问报告预计于明年4月完成。

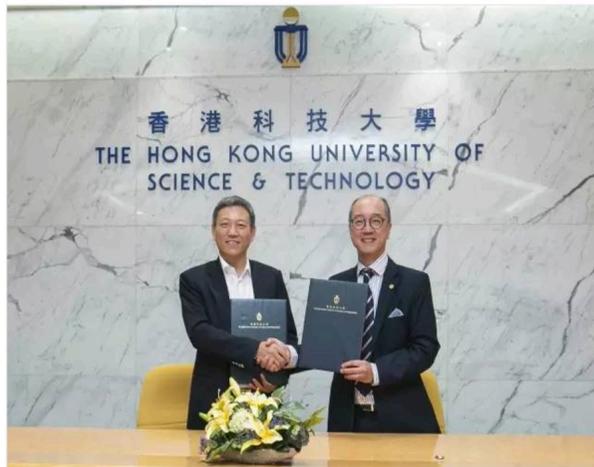
29 June 2016
People.cn

Publicity & Media coverage

Digital China, 11 Sep 2015

神州数码牵手港科大，共建智慧城市（香港）研究院

(Original) 2015-09-11 神州数码控股 神州控股



协议约定，围绕双方的资源优势和战略发展规划，港科大与神州数码将在智慧城市发展规划、市民服务、政府决策建议、大数据处理、云计算、高可靠软件、嵌入式与普适计算、车联网、未来媒体、未来网络、遥感数据处理、智慧医疗、智能交通、物联网信息技术等领域重点开展合作。双方将加大投入研究人员（教授和研究生）、配套资金、设备、实验环境等资源，推动在智慧城市标准、专利、产品技术等多方面取得创新性的成果。双方将成立联合管理委员会、技术委员会、项目联络办公室等机构，保障合作的顺利推进。



2015年9月10日，神州数码控股有限公司与香港科技大学签署合作框架协议，双方将共同成立智慧城市（香港）研究院，并基于长期合作伙伴关系，展开智慧城市领域的深度合作。神州数码控股董事局主席郭为与香港科技大学校长陈繁昌教授代表双方签署协议。

QQ.com, 19 Dec 2015

专访微信人工智能实验室：社交网络已成为人工智能的一个热点

作者 杜小芳 发布于 2015年12月19日。估计阅读时间：不到一分钟 | 讨论

分享到： 微博 微信 Facebook Twitter 有道云笔记 邮件分享

腾讯旗下微信团队和香港科技大学于11月30日宣布成立联合实验室，WeChat-HKUST Joint Lab on Artificial Intelligence Technology，简称：WHAT Lab。该实验室将以人工智能为主要研究方向，旨在改善用户的生活服务体验，借助大数据拓展机器学习的边界。

背景：

微信是腾讯于2011年推出的一款手机社交媒体应用，与开放性的舆论平台不同，微信是一种封闭的社交平台。发布之后得到了迅猛的发展，到现在它拥有了6.5亿的活跃用户，创造了中国IT的又一个奇迹。



THANK YOU FOR WATCHING

BDI Website: <http://bdi.ust.hk>

About BDI

Leveraging on the existing strength and resources of the Hong Kong University of Science and Technology (HKUST), the Big Data Institute (BDI) was established upon the needs from both industry and society for a new model for managing multi-disciplinary focal points for Big Data research, by coordinating faculty's research on Big Data and Data Science at HKUST. BDI will provide a strong and highly visible leadership role in Big Data and Data Science research in Hong Kong and the world...

Campus Location



Contact Us

Email: bdi@ust.hk
Website: <http://bdi.ust.hk>
Webmaster: ycaiap@connect.ust.hk



香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY