産品可靠性暨系統安全研發中心
CENTRE FOR ADVANCES IN RELIABILITY AND SAFETY
A Research Centre admitted under *AIR@InnoHK* Cluster

FEBRUARY 2022 ISSUE 5

CAiRS E-NEWSLETTER

▶▶▶▶▶▶▶▶▶ ■ ■ ■ ■

Dr Vincent Ng
Senior Scientific Officer, University Research Facility in Big Data Analytics (UBDA)
The Hong Kong Polytechnic University
PhD (Simon Fraser), MMath (Waterloo)
Visiting Research Fellow in CAiRS
吳道義博士
香港理工大學大數據分析中心實驗室高級科學主任
西門菲莎大學哲學博士，滑鐵盧大學電子計算數學碩士
產品可靠性暨系統安全研發中心客席研究員

Reliability Analysis through Smart Big Data

運用智能大數據分析可靠性

*: A Conversation with Dr Vincent Ng
與吳道義博士對話

Centre for Advances in Reliability and Safety (CAiRS) links academic and industrial counterparts to introduce and implement artificial intelligence methods and machine learning techniques with reliability modelling, with the goal of improving product reliability and system safety. Data analysis is one of the important areas in the research. In this issue, we invited Dr Vincent Ng who is a Senior Scientific Officer at the University Research Facility in Big Data Analytics (UBDA) of The Hong Kong Polytechnic University to share his ideas on smart big data analysis. Dr Ng has rich experience in big data analytics, machine learning, and medical informatics. Also, he joined CAiRS as a Visiting Research Fellow after his retirement from Associate Head (Teaching) of the Department of Computing in The Hong Kong Polytechnic University (PolyU).

產品可靠性暨系統安全研發中心（CAiRS）與學術界及業界夥伴共同合作，引入及推行具備可靠性建模的各種人工智能方法及機器學習技術，目的是改善產品的可靠性和系統安全。對研究來說，數據分析是其中一個不可或缺的步驟。我們在今期邀請到香港理工大學（理大）大數據分析中心實驗室（UBDA）高級科學主任吳道義博士，分享一下他對智能大數據分析方面的見解。吳博士在大數據分析、機器學習，以及醫療資訊學上擁有豐富的經驗，並且自理大電子計算學系副系主任（教學）退休後，以客席研究員身份加入CAiRS。

Research Focus

: **What is your research interest?**
你的研究興趣是什麼?

My research topic focuses on data analysis. After graduation from Canada, I spent over ten years on bioinformatics data analysis in a cancer research institute including image analysis and epidemiological studies. Since I joined PolyU in 1994, I started my research topics on data mining and artificial intelligence with different projects on system design and data-driven modelling. I also developed projects related to education across different age ranges of students which analyzed to get insight not only into academic performance but also the emotional and career development of students.

我的研究主題以數據分析為主。在加拿大畢業後,我在癌症研究所從事超過十年有關生物資訊學數據分析的工作,包括成像分析及流行病學研究。自1994年加入理大後,我參與多個系統設計和數據建模項目,並展開有關數據挖掘和人工智能方面的研究。此外,我亦就不同年齡組別的學生開發了教育相關項目,透過分析不僅能了解學生的學術表現,亦能探討他們在情緒和職業生涯方面的發展。

: **Do you have any suggestions on using big data for reliability analysis?**
對於使用大數據來進行可靠性分析,你有任何建議嗎?

Nowadays, there are lots of data around us with different acquisition channels. In the big data world, there is an important aspect, smart data. We are now considering not only the volume of data but also the variety and velocity of data. If the data is smart, we can conduct meaningful analyses on feature discovery and get insights through the relationships among the data. This concept of effective and smart data can also apply to reliability analysis.

今天,我們可以透過不同的渠道獲取大量數據。在大數據的世界裡,智能數據相當重要。目前我們不僅注重數據的數量,亦需要重視數據的多樣性和速度。如果擁有智能數據,我們便能具針對性地尋找特徵並進行具意義的分析,以及透過數據之間的關係獲得相關資訊的深刻見解。這個有效且智能的數據概念,亦適用於進行可靠性分析。

Research Focus

: Would you elaborate more about what is smart data?
你可否詳細解釋一下什麼是智能數據?

When we consider smart data, we shall focus on the quality of data. It is good if the data are continuous and holistic. We generally prefer high completeness of data without any gaps and missing information. This is to prevent non-reliable results on the analysis. The other consideration is the age of data. If the data is too old, the scenarios and domains may be different. Hence, we should work with the related experts with rich domain knowledge for application validation. Furthermore, the cost of data collection including human resources, system development, and detailed planning for the data collection is important at the beginning of the project. We shall also carefully clarify how and what data shall be collected before we execute the process.

當我們提及智能數據時,我們應以數據的質量為主,而較理想的是擁有延續且整體的數據。一般而言,完整度高,並且不存在任何間斷和沒有遺漏的數據較為可取,這樣可避免對分析造成無法預測可靠的結果。另一個考慮因素是數據的「年齡」,假如數據太舊,那麼涉及的情況和領域都有可能在現時進行的分析有所入。因此,我們應與在相關領域中專家合作,令數據校驗以可效地應用數據分析。如人力資源、軟件系統開發,這些分析項目需要有詳細規劃的收集數據與估計附帶的成本,及項目初期亦必須謹慎考慮。在開始收集數據之前,我們應仔細訂明收集方法以及收集哪些數據。

有效性 Validity	數據是否能夠按照預期達到量度的目的 Data measure what they are supposed to measure
可靠性 Reliability	每個人時刻都以相同的方法定義、量度和收集數據 Everyone defines, measures, and collects data the same way — all the time
完整性 Completeness	數據包含了所有需要用作計算指標的數值 沒有遺漏任何變量 Data include all the values needed to calculate indicators No variables are missing
精準度 Precision	數據具有充足的細節;量度單位非常清晰 Data have sufficient detail. Units of measurement are very clear
合時性 Timeliness	使用最新的數據;資料可合時地提供 Data are up to date. Information is available on time
真確性 Integrity	數據真確無誤。數值並無出現任何故意造成的偏差,亦無因政治或個人理由而被更改 Data are true. The values are safe from deliberate bias and have not been changed for political or personal reasons

Figure 1: Elements of Data Quality [1]   圖 1:數據品質的基本要素 [1]

[1] Adapted from: http://www.cpc.unc.edu/measure Presentation by Win Brown, USAID/South Africa, School of Health Systems and Public Health, Monitoring and evaluation of HIV/AIDS Programs, Data Quality; March 2, 2011
擷取自:http://www.cpc.unc.edu/measure(Win Brown 的報告演講),USAID/South Africa, School of Health Systems and Public Health, Monitoring and evaluation of HIV/AIDS Programs, Data Quality(USAID/南非,醫療系統及公共衛生學院,愛滋病毒／愛滋病項目的監測及評估,數據品質);2011年3月2日

Research Focus

: One of the approaches of reliability analysis is data-driven modelling, would you share your thought on it?
可靠性分析的其中一個方法是數據驅動模型，
可否分享一下你對這方面的意見?

I would like to share two ideas about the data-driven modelling. The first one is data can be used as evidence to assist the decision-making process. We can set up the hypothesis and conduct pilot experiments for this purpose. The other one is feature engineering and discovery. We can analyze the relationships across different data sources and discover anomalies or new insights which we are not aware of before. Furthermore, predictive modelling can be the next step based on these features. For data collection, we can start by asking whether we have enough data, and then consider acquiring data, labeling data and how to improve the existing data, labels and models.

我想與大家分享兩個有關由數據驅動的模型之想法：第一個想法是數據可用來作為憑證以協助我們進行策，並據此設定假設並進行試行實驗；另一個想法是用尋找及對特徵進行數據工程，當中我們可以分析出不同數據源之間的關係，讓我們對過往不認知的異常情況有更深入的了解或產生新的想法。另外，我們可基於這些特徵進行下一步的預測建模。至於數據收集方面，我們可以先思考一下自己是否擁有足夠的數據，然後考慮如何獲取和標籤數據，以及如何改善現有的數據、標籤方式和運作模型。
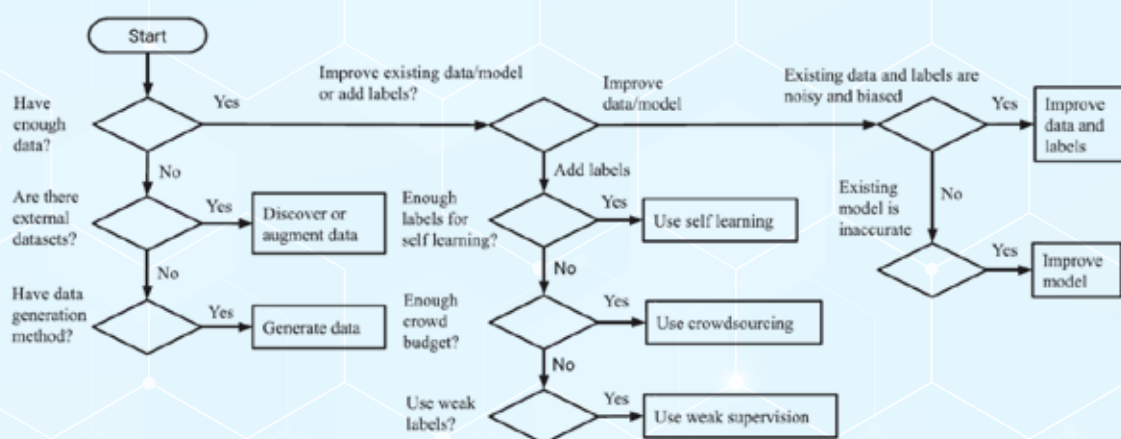

Figure 2: A decision flow chart for data collection [2]　　圖 2：數據收集的決策流程圖 [2]

[2]: Yuji Roh, Geon Heo, and Steven Euijong Whang, "A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective", IEEE transactions on knowledge and data engineering, 2021-04-01, Vol.33, April 2021
Yuji Roh、Geon Heo 及 Steven Euijong Whang，《A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective（機器學習的數據收集調查：大數據 – 人工智能集成的想法）》，IEEE transactions on knowledge and data engineering（知識及數據工程的IEEE學報），2021-04-01，第33期，2021年 4 月

: Would you share some considerations on anomaly analysis?
對於異常情況的分析，你認為我們應注意哪些事項?

In our projects, outliers or anomalies are the common elements to be identified during the analysis. We shall be taken note that the anomalies can occur in different contexts and environments. Take environmental temperature as an example, it can be an anomaly if the measured temperature is as high as around forty degrees in a general office, but it may not be classified as an anomaly in the kitchen especially busy working hours. This simple illustration indicates that we need to understand the context, especially the domain knowledge on data labelling. In real applications, it is quite common that there is unbalanced data between normal and abnormal information. The proportion of abnormal data may be as low as single-digit percentage. Therefore, we utilize good techniques to get more significant information from feature engineering and sampling techniques.

在我們的項目中，離群值或異常情況經常都會在分析期間出現，我們應留意異常可以在不同的景況及環境中發生。就以環境溫度為例，假如一個普通辦公室中，量度出來的溫度高達大約四十度，我們可視之為異常；但是如果在廚房（尤其是營業時間內）中量度出同一溫度，則不應歸類為異常。透過這個簡單的例子，我們明白到在使用數據時首先我們需要了解當中的景況，尤其是為數據背景進行標籤的技術。對於實際的應用環境裡，正常和異常數據往往在數量上會出現不平衡的情況，而出現異常數據的百分比比例更可能低至個位數字。因此，我們會運用各種技術並透過對特徵工程來獲得更具意義的資料。

News Highlights

PolyU and University of Maryland from USA, jointly established the
Centre for Advances in Reliability and Safety (CAiRS)
Combining AI and innovative technologies to improve product safety and system reliability

理大聯同美國馬里蘭大學(UMD)共同成立
產品可靠性暨系統安全研發中心 (CAiRS)
揉合 AI 及創新技術提升產品安全及系統可靠性

In our daily life, there are many reliability and safety issues. Electronics degrade due to complex electronics ageing, latent software faults, and the interactions between the two. Also, electronic system failures are inevitable because of the current methods to assess reliability and safety. These issues are very likely to lead to serious consequences.

In view of this, two universities, The Hong Kong Polytechnic University (PolyU) and the University of Maryland - College Park (UMD) from USA, have jointly established a research center, namely - Centre for Advances in Reliability and Safety (CAiRS). The Centre gathers top local and overseas researchers, uses advanced equipment and leverages innovative artificial intelligence technology in order to conduct wide ranging product reliability and system safety research to accurately predict the occurrence of failures and prevent them from occurring.

CAiRS has been admitted as one of the research centres in the InnoHK Clusters, a major initiative of the HKSAR. CAiRS has carried out five research programmes to date, namely "Anomaly Detection and Syndromic Surveillances", "Innovative Diagnostics for Health Management", "Prognostics for Remaining Useful Life Assessment", "Safety Assurance: Improve functional safety" and "Data Analytics Platform for Reliability". Under the 5 projects, there are 15 projects running in parallel. The range of applications of the research is very wide, including robots, medical devices, vehicles, telecommunications, consumer goods, public utilities, transportation, microelectronics, electrical installations, sensors, IoT products and advanced manufacturing. Moreover, CAiRS has signed NDA with 31 well-known and representative local companies from the industry to kick start research collaboration. At present, we have over 20 collaborative projects already running in progress to improve reliability and safety of their products and systems.

很多日常生活用品，都會因為電子或機械零件老化、軟件故障或環境因素等等，而出現可靠性和安全性問題，簡單的可能是電線短路，嚴重的卻可能導致嚴重的後果，例如機械過熱導致火警等。有見及此，香港理工大學(理大)及美國馬里蘭大學帕克分校(UMD) 攜手合作成立研發中心 ─ 產品可靠性暨系統安全研發中心(CAiRS)，匯聚全球頂尖科研人員，裝置了先進的儀器設備，配合創新人工智能技術，以進行各項產品可靠性和系統安全研究，務求準確預測故障發生，防患於未然。

CAiRS早前獲選進駐由特區政府的旗艦項目「InnoHK創新香港研發平台」，並正就五個範疇展開研究，包括異常檢測與症狀監測、產品健康管理的創新診斷、剩餘使用期限預測、安全保障：提升功能安全性，以及數據可靠性，共15個項目。這些項目應用範圍非常廣泛，包括機械人、醫療器材、車輛、電訊、消費品、公共事業、運輸、微電子、電力裝置、傳感器、物聯網產品及其他與先進製造相關的範疇。並已與三十一家本地知名企業簽定協議籌備研究項目。現時，我們有超過二十個研究項目已進行中，旨為業界提升產品和系統的可靠及安全。

理大一直因應業界及社會需要，進行跨學科研究，當中涉及工程，科學及前沿技術。CAiRS匯集了理大工程學院及 UMD

News Highlights

PolyU is committed to conducting state-of-the-art interdisciplinary research in response to the needs of industry and society. With over 20 academics and scholars from the Faculty of Engineering of PolyU and UMD, having research excellence and track records in product reliability, and with strong support from industry, CAiRS will bring significant benefits and contributions to smart city development and smart manufacturing.

超過二十位學者專家，加上 UMD 在產品可靠性的卓越研究基礎，及業界的支持，將就智慧城市及先進製造兩大方面作出貢獻。

For more details, please scan the QR code
請掃瞄二維碼查看

Group Photo (from left)
Prof. KM Lam, Prof. HC Man, Dr. CH Ng,
Prof. Wing-tak WONG & Ir. Prof. Winco Yung

（左起）
林建文教授、文効忠教授、吳自豪博士、黃永德教授及容錦泉教授合照

Prof. Wing-tak WONG, Deputy President and Provost of The Hong Kong Polytechnic University
香港理工大學常務及學務副校長黃永德教授致辭

Prof. Wing-tak WONG, Deputy President and Provost of PolyU, said, "Harnessing its advanced equipment and, top-notch scientific research talents, CAiRS is dedicated to the research and development of breakthrough technologies. Their research solutions can be widely adopted by all industries in Hong Kong that value reliability and system safety. PolyU has been actively cooperating with world-renowned universities and establishing close partnerships with industry to benefit society through cutting-edge research. I believe that CAiRS can effectively translate scientific research results to real world solutions, creating positive impact for various industries as well as society."

香港理工大學常務及學務副校長黃永德教授指：「CAiRS 擁有先進設備，頂尖科研人才，專注研發突破性的技術，研究範圍更廣泛應用於香港所有重視可靠性及系統安全的行業。理大一直積極透過與世界知名的學府合作，並與業界建立緊密伙伴關係，為相關的研究創造突破，成就具影響力的科研。我相信 CAiRS 能夠為業界提供良方妙策，有效將科研成果轉化為具體的解決方案，為社會帶來裨益及貢獻。」

News Highlights

Dr. CH Ng, Chairman, The Hong Kong Electronic Industries Association said, "The establishment of CAiRS a platform for global research collaboration is welcomed by industry. CAiRS uses artificial intelligence for reliability and safety analysis, and works with industry to promote the development of products or brands, ultimately, helping Hong Kong to become a highly respected and trusted 'international brand'. I believe that industry and CAiRS will continue to cooperate seamlessly in the future to make Hong Kong a key R&D centre with international authority."

香港電子業商會會長吳自豪博士表示：「業界非常樂見CAiRS的成立，並成為匯聚眾多科研專才的平台。CAiRS透過運用人工智能進行可靠性和安全性分析，與業界一同推進產品或品牌的發展，幫助香港成為備受推崇和信賴的『國際品牌』。深信未來業界與CAiRS會繼續合作無間，令香港成為具國際權威的重點研發中心。」

**Dr. CH Ng, Chairman,
The Hong Kong Electronic Industries Association
香港電子業商會會長吳自豪博士致辭**

**Ir. Prof. Winco Yung, Centre Director and Executive Director,
Centre for Advances in Reliability and Safety (CAiRS)
產品可靠性暨系統安全研發中心(CAiRS)
總監及執行董事 容錦泉教授介紹 CAiRS**

Ir. Prof. Winco Yung, Centre Director and Executive Director, Centre for Advances in Reliability and Safety (CAiRS) said, "CAiRS focuses on the use of artificial intelligence to develop new personalized management models. The application and results of the Centre's research are very important to the development of smart cities. The scientific research team of CAiRS and I are delighted to use our expertise to collaborate with partners in different industries. CAiRS will build an international brand for the products and systems design, made and commissioned in Hong Kong, and contribute to the development of smart city and smart manufacturing."

產品可靠性暨系統安全研發中心(CAiRS)總監及執行董事容錦泉教授表示：「CAiRS 專注利用人工智能研發新型個性化管理模式，研究的應用及成果對智慧城市的發展十分重要。我和CAiRS 的科研團隊非常高興能運用我們的專業，與不同行業合作夥伴協作，將香港設計、製造的產品或系統打造成國際品牌之餘，亦為建設明日香港成為智慧城市及先進製造中心作出貢獻。」
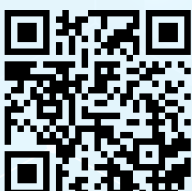
News Highlights

: **Webinar on
"Advances in Predictive Maintenance and
System Safety Assessment"**

We organized a webinar on 15 December 2021. For more information, please visit our website at www.cairs.hk

我們在2021年12月15日舉行了網上研討會。詳情請到CAiRS網址 www.cairs.hk

For more details, please scan the QR code
請掃瞄二維碼查看詳細資料

News Highlights

: CAiRS participated in Hong Kong Electric (HEC) Professional Workshop
研究團隊參與港燈專業工作坊

CAiRS' project team participated in professional workshop at Cable Jointing Training Centre of HEC
CAiRS團隊到港燈培訓中心參與專業工作坊。

Hong Kong Electric (HEC) is one of CAiRS' collaborators. Our research project with HEC is the study of reliability and health index of the underground cable. HEC specially arranged a professional workshop for the CAiRS research team on January 4, 2022.

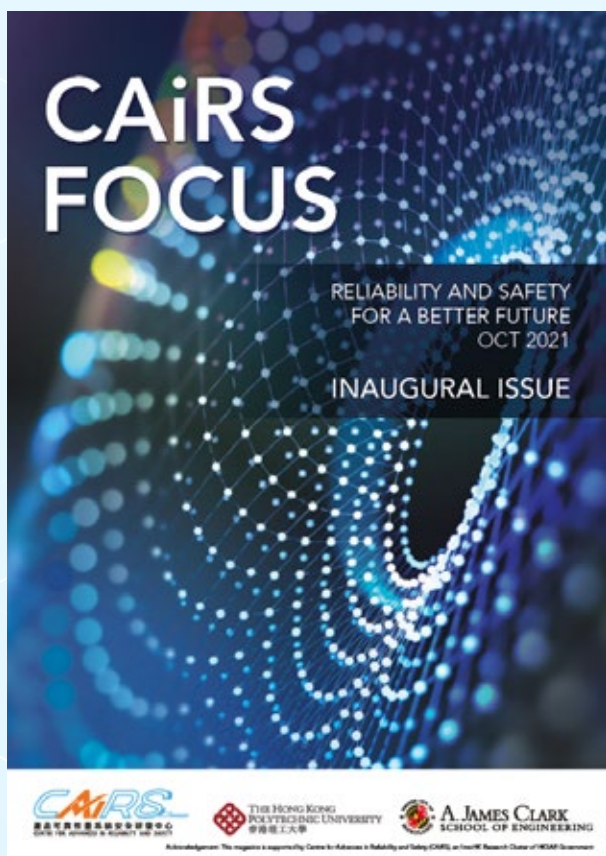港燈(HEC)是我們其中一家合作伙伴，我們與HEC的研究項目是研究地下電纜的健康指數和可靠性。最近2022年1月4日，港燈為CAiRS研究團隊安排了一次專業的工作坊。
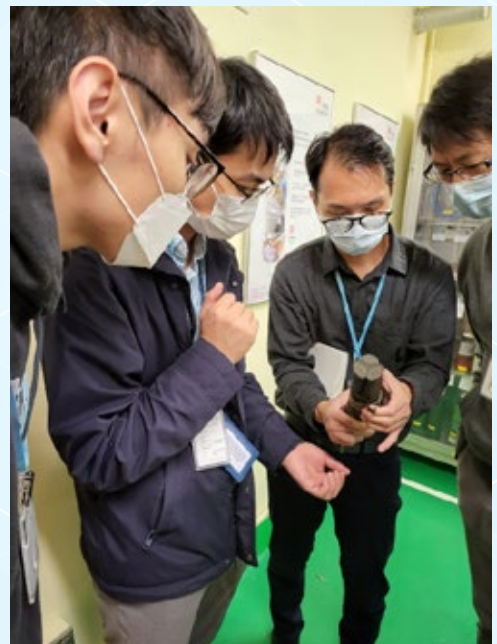
CAiR's Project team examined the underground cable.
CAiRS研究團隊檢查地底電纜。

Publication

: CAiRS Focus
CAiRS 年刊

We published the inaugural issue "CAiRS Focus", which provides the latest information and research projects preview of CAiRS, a research centre for product reliability and system safety. CAiRS Focus will be sent to our Collaborators, Industries Association, Professional organization, Institutions and Government departments. You're welcome to view the e-copy of CAiRS Focus at our website www.cairs.hk. We will update the latest information about product reliability and system safety from time to time. Please feel free to contact us if you have any questions on our publication.

為了讓大家對CAiRS的研究項目有更深入了解，我們每年會製作年刊以供大家參考，創刊號經已出版，亦可於CAiRS網址瀏覽。CAiRS Focus主要派送到合作伙伴、商會、專業機構、大專院校及政府部門等。我們會繼續為大家介紹最新關於產品可靠及系統安全的資訊。如有任何有關刊物的查詢，歡迎與我們聯絡。

To read the publication, please scan the QR code
請掃瞄二維碼閱讀