広島工業大学紀要教育編
第24巻（2025）65-74
報　　　　告

理工系学生用の英文で学ぶ最新データサイエンス教材

釜江　常好[*]・鈴木　貴[**]

（令和６年10月31日受付）

Materials to Learn Cutting-Edge Data Science in English for Students Science in Science and Engineering

Tsuneyoshi KAMAE[*] and Takashi SUZUKI[**]

（Received October 31, 2024）

Abstract

To prepare for diverse career paths, students have to learn diverse fields of science and engineering. Education a student receives in a Japanese university, however, is limited to what the faculty and the department he/she belongs to offer. The number of instructors belonging to one faculty or one department is quite limited in most Japanese universities, when compared to those in the US or in Europe. This shortcoming has been preventing students to learn and be trained in emerging new fields in the neighboring fields.

Among leading universities there is another implicit pressure among students studying mathematics and physics that the “best” mathematics and physics take the “pure and simple” forms. This way of thinking has been formed in the earlier half of the 20th century when several revolutionary ideas have been proposed. For example, the 4-color theorem in mathematics was proven by the team who had better access to multiple IBM-370, the super-computers in mid 1970s than competitors[1]. The dream proof of Fermer’s theorem took hundreds of pages. (1: Robin Wilson: Illinois Journal of Mathematics Vol 60 (2016), pp 149-178)

In particle physics the super-string theory has been attracting many talented scientists worldwide. No body yet sees the ultimate theory.

Meanwhile, fields including astrophysics, medicine, molecular and biological sciences have become to rely heavily on data analyses. The data come either from observations, measurements, or simulations.

Many developments in data science came out of physics: the famous Fermi-Pasta-Ulam note led to the theory of solitons, the KAM theory and machine learning algorithms. Feynman-Kac formalism is based on the path-integral in quantum mechanics. It finds hundreds application in the fintech field. Sadly, many physicists and mathematicians do not know about these developments.

The teaching material have been prepared to catch attention of students and instructors in the fields of science.

[*] Emeritus: University of Tokyo (Physics) and Stanford University (SLAC and KIPAC)
[**] Hiroshima Institute of Technology, Electronics and Computer Engineering

Acknowledgements

TK acknowledges assistances given by Drs. M. Oguri（Chiba U.), A. Ukawa（Tsukuba U.), and T. Yamagata（U. of Tokyo).

Table of Contents

Contents of Slides（1/

	Title	Item	Contents
1	Make Best Use of your PC 2024	a	Know you PC’s specifications (Windows 10 and 11)
		b	Check that Hyper-V is enabled
		c	Download x64 or x86 depending on your PC architecture
		d	Use WSL 2 because it runs a real Linux kernel, through a subset of Hyper-V features.
		e	Install Ubuntu on WSL2. Download from Microsoft store:
		f	Setup your super-user password (used when running >sudo)
		g	Access the Windows C: file system by change directory cd /mnt/c/users/your user id/
		h	Install applications using >sudo apt install, >sudo apt update, >sudo wget, >sudo curl
		i	>sudo install python3.10 or >sudo install python3.11
2	Intro to Astrophysical Data Analyses	a	Fits viewer fv
		b	Download data of Fermi Large Area Telescope γ-ray observatory
		c	Sample Astronomical Data Analysis: Fermi-LAT + DS9
		d	Sample Data Analysis: Orion mol cloud (Fermi-LAT and DS9)
		e	View images taken by Subaru SupremeCam with Aladin
		f	Crab Nebula images by Chandra (X-ray) with Aladin Sky Atlas
		g	Make intensity contours from the Chandra (X-ray) image
		h	Review for Reports on Progress in Physics: Crab pulsar
		i	Andromeda Galaxy (M31) B-filter image at Digitized Sky Survey
		j	Make a constellation map and convert to Boat-Shaped image

Contents of Slides（2/

	Title	Item	Contents
3	Strong Gravitational Lensing	a	A complete Einstein Ring by James Webb Space Telescope
		b	Einstein Cross (from Hubble and WIYN)
		c	Complex distorted grav. lensing image incl. a ring
		d	Lensing by a large galaxy cluster GAL-CLUS-022058s
		e	Complex Gravitational Lensing by a dense cluster of galaxies
		f	HST Legacy: mos_0441167 ACS/WFC F555W (mosaic)
		g	Strong gravitational lensing of explosive transients
4	JWST Science Data Analysis		Do no download Data using Bulk Downloads: it takes forever. Simple way No.1 is to download JWST images in jpeg and edit them.
			Simple way No.2 is to download the first release data. Set“Release Date”btwn 2022-07-13 14:00 and 2022-07-013 16:00 and download. Find “filter selection” and choose f444, f200, f335.
			Open the 3 fits files in DS9 to make a combined false color image Read in the 3 fits files one by one and assign an RGB color, RGB.
5	Transform Cartesian to Zenith Azimuth Map	a	Install ImageMagick and convert cartesian Earth map and cartesian Fermi Lat All Sky maps to plots Zenith Azimuth Maps.
6	Planar Map to Spherical Map	a	Download Universal_Transverse_Mercator_coordinate_system from Wikipedia. The zoning is at a regular longitude interval (6 deg).
		b	Calculate the two longitudinal boundaries of a UTM zones.
		C	Use python codes provided and boat-shaped swaths and cuts.
7	QGIS Shape File to CSV and Excel		Learn about QGIS, the geographic information system (GIS) software
			Install QGIS program to your Windows 10/11
			Import shape (*.shp) files from QGIS Desktop: ne_110m_coastline.shp
			Simplify or smooth shape files
			export OS Boundary Line shape (*.shp) files to csv's containing the boundary line in latitude/longitude

Contents of Slides（3/

	Title	Item	Contents
8	General Statistics and Inference	a	Data analysis: the process of using data analysis to infer properties of an underlying distribution of probability.
		b	Descriptions of statistical models: 3 levels of modeling assumptions. Fully parametric; Semi-parametric Nonparametric
		c	No one-size-fits-all method: Residual plots and Cross validation
		d	Paradigms for inference: Frequentist and Bayesian inferences
		e	● Likelihood-based inference: ● Assume a specific set of parameter values (θ’ s) ● Formulate the statistical model such as normal distribution ● Construct the likelihood function (with parameters θ’ s) ● Maximizing the likelihood function ● Assessing adequacy of the model and assess uncertainty ● Inference and interpretation
		f	Glossary of probability and statistics; acronyms used in probability and statistics; Statistical distributions;
		G	Distribution Density Functions (plots and python codes): and Simple and Metropolis Rejection Sampling (python codes and plots)
9	Birth of Numerical Experiments (Monte Carlo) in Science	a	Fermi-Pasta-Ulam-Tsingu paradox : They added a weak nonlinear interaction and found that almost all energy in an oscillator system come back to the lowest frequency => the FPU of FPUT paradox.
		b	Pursuit for the solution of the FPU paradox led to soliton science, chaos science and the KAM theory.
		c	Discovery of solitons in Korteweg-de Vries (KdV) equation in a computer exp by N. J. Zabusky and M. D. Kruskal
		d	KAM theory and stability all kinds of orbiting motion: Stability in solar system, Saturn’ s ring, asteroid distribution in the solar system.
10	Bayes Inference VS Frequentist’ s Inference	a	Bayes Theorem assumes no prior. The theorem is stated in many different ways in the literature and terminologies are confusing.
		b	Events E in a sample space S: A be any event in S. Then conditional probability $P(E_i \mid A)$ means prob of E_i given that A is observed
		c	Statement may be written differently. Events $\{E_i\}$: the hypothesis. $P(E_i)$: the prior prob of hypothesis $\{E_i\}$; $P(E_i \mid A)$: the posterior prob of hypothesis $\{E_i\}$; $P(A \mid E$i): likelihood A occurs under hypothesis {Ei}.
		d	A graph is plotted to understand the concept above.
		e	Bayes theorem works well when there are hidden (latent) variables: Bayes’ billiard table arguments (referred to a python codes)

Contents of Slides（4/

	Title	Item	Contents
11	Pseudo Random Number Generator and Mersenne Twister	a	History of Pseudorandom Number Generators
		b	Multiplicative congruential generator (MCG)
		c	LFSR extension, generalized feedback shift register (GFSR)
		d	Matsumoto and Kurita (1992) introduced a variant called the twisted GFSR (TGFSR)
		e	Mersenne Twister by Matsumoto and Nishimura (1998) ● It is based on the Mersenne prime number 2ˆ(19,937) — 1 ● It can be implemented using a 32bit word length. ● It is often called the MT19937-32 generator.
		f	Many new algorithms are proposed yearly
		g	There are a few Assessment Test suits for RNGs are available. NIST's Stat Test Suite for Pseudo RNGs (2001)
		h	Explanation of Mersenne Twister
12	Randomness test suite master by NIST	a	NIST's Statistical Test Suite for Pseudorandom Number Generators – Python implementation –
		b	● Test For Frequency Within A Block ● Runs Test ● Test For The Longest Run Of Ones In A Block ● Random Binary Matrix Rank Test ● Discrete Fourier Transform (Spectral) Test ● Non-Overlapping (Aperiodic) Template Matching Test ● Overlapping (Periodic) Template Matching Test ● Maurer's Universal Statistical Test ● Linear Complexity Test ● Serial Test ● Approximate Entropy Test ● Cumulative Sum (Cusum) Test ● Random Excursions Test ● Random Excursions Variant Test
13	Expectation Maximizer with Faithful Geyser and Iris Dataset	a	K-Means Clustering gives for a given number of clusters. Elbow point and Silhouette score give the best number of clusters
		b	Expectation-Maximization Algorithms: Gaussian Mixture Model
		c	Approximate nearest neighbors in t-distributed stochastic neighbor embedding (TSNE) applied to the hand-written numeral dataset,

Contents of Slides（5/

	Title	Item	Contents
14	Earth Orbit Eccentricity, Rotation Axis and Milankovitch Model	a	Earth’s Milankovitch Cycles: the long-term, collective effects of changes in Earth’s position relative to the Sun are a strong driver of Earth’s long-term climate, Milankovich Cycles.
		b	Axial obliquity (tilt)ε , Eccentricity of the orbit (*e*), Daily-average insolation at top of atmosphere on summer solstice at 65N.
		c	Data fitting to temperature models of late Pleistocene deglaciations:
		d	The 100kyr cycle and 44kyr cycle reproduced by the model.
		e	Change in eccentricity, obliquity, precession of the Earth orbit and precession is likely influenced the birth of Homo Sapiens.
15	Markov Chain Monte Carlo: History and Review	a	History leading to proposal of the Monte-Carlo Method ● Calculation of the chain reaction in enriched U-235 in UK during WWII (MAUD Committee) ● Serber-Wilson method during WWII (US DoE LA-1391) ● Diffusion eqn was solved numerically byT-5 desktop calculators mostly by wives of Los Alamos scientists. ● Paper by A.A. Markov (1906) has not been explicitly cited in publications on the Monte Carlo method. ● Nicholas Metropolis, S. Ulam: The Monte Carlo Method (1949) ● N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller; J. Chem. Phys. 21, 1087 (1953)
		b	Sampling needs to be efficient: Metropolis-Hastings algorithm,
		c	Path to the second-generation MCMC revolution ● Gibbs sampler (Geman, Geman,1984) ● Hamiltonian (Hybrid) Monte Carlo introduced by Duane et al. for lattice field theory (1987) and by B.Mehlig, et al, for molecular dynamics (1992). Desirable features for MCMC, the time-reversal and detailed balance come out naturally ● The leap-frog method has been adopted in 2003 ● No-U-Turn HMC introduced to improve sampling efficiency
		d	Hamiltonian Monte Carlo method applied to coseismic fault model estimation by Yamada, Ohno and Ohta in 2022
		e	Comparison of convergence speeds: random-walk MH vs HMC
		f	Possible artificial correlation: No-U-Turn, HMC, Gibbs samplers
		g	Unwanted correlation in sampling: Adaptive Metropolis Algorithm
		h	Improve convergence: Delayed Rejection + Adaptive Metropolis => DRAM inference

Contents of Slides（6/

	Title	Item	Contents
16	Sequential Monte Carlo and Feynman-Kac Formula	a	Hamilton MC vs Sequential MC ● HMC performs well with continuous distri with ″weird″ shapes ● Sequential Monte Carlo methods—also known as particle filters—offer approximate solutions to nonlinear state-space systems
		b	The particle filters may be interpreted as mean-field particle interpretations of Feynman-Kac probability measures.
		c	Importance sampling is essential. It find where the current estimation deviates from the target and enhances that part for next sampling.
		d	Feynman-Kac formula establishes a link between parabolic partial diff eqns and stochastic processes. It finds applications in finance
		e	Feynman-Kac formula provides an integral representation for the solution and hence helps to establish error estimates.
17	Self-Organizing Maps and Hopfield Network	a	Two roots led to the Self-organizing maps: one comes from Teuvo Kohonen (1982) and the other comes from J. Hopfield (1982, 1984).
		b	Kohonen map: sort objects (eg. MINST) by shapes and group them. Hopfield network fills in missing parts or remove noises (eg. MINST). Because Hopfield net is single-layer, it can be trapped in local min.
		C	Demonstration python codes: ● Traveling Salesman Problem ● Clustering using shape (MIST) ● Correct shapes to match the example
18	Raspberry Shake and Anatomy of the Earth	a	
		b	
		c	
		d	
		e	
19	Thermo-Vacuum Former of Tactile Sheets to Enhance Accessibility to Data Science	a	A complete set of drawings to make an inexpensive thermo-vacuum former do-it-yourself way. (Material cost less than 200USD).
		b	Decompressor: Thin vacuum chuck; Vacuum box; High power vacuum cleaner
		c	Heating plate assembly: Aluminum heat reservoir
		d	Photos showing working procedure, step-by-step, to transfer tactile graphics to A4 size Brailon sheets.

Appendices

報告書 1 電子情報工学科　今岡剛人

・行ったこと

- ・DS 9 を使った宇宙データの観察
- ・ImageMagick を使った画像加工
- ・QGIS を使った世界地図の作成
- ・QGIS を使ったハザードマップの制作

・学んだこと

私は今回の活動を通し、 データの解析、一般的なことから専門的なことまでをフリーソフトを使って行うことができることを知った。また fits ファイルや gis などの様々な形式のデータファイルに初めて触れ、それらが国内外の多くの機関から提供されていることを知った。提供されているデータで必要なものを用いることで新たにデータを作る方法を学んだ。

・今後の課題

今回私は QGIS をメインに活動を行ってきたが、まだデータを張り付けるくらいのことしかできていなかったため今後は使用するデータに手を加えてもっと目的にあったデータを作れるようにしたい。具体的には今回のハザードマップでいうと「浸水の深さが分かるようにする」ということができるようにしたい。

・感想

この活動をしていくなかでいままで触れてこなかった分野のデータやソフトに触れることができた。今まで興味を持っていなかったことに興味をもち、自分自身の知識の幅を増やすことができたと思う。特に今回私が使った QGIS は用いるデータによって本当にたくさんのことができるため使えるようにしておくと便利だと感じた。先日のミーティングでもお話があったが、業務にも使用することができるためもっとできることを増やしたい。

ここまでの活動は自分の視野や知識の幅を広げるいい機会になったと思う。

報告書 2 電子情報工学科　田村　喬

RISC と ARM について

私は RISC と ARM についてとその関係を調べました．

RISC は CPU アーキテクチャの一種で，簡単な命令を高速に複数処理することでパフォーマンスを向上させるアーキテクチャでした．

また，RISC と対になるアーキテクチャとして CISC がありました．CISC は複雑な処理をできるだけ少ない命令回数で処理することでパフォーマンスを向上させるアーキテクチャでした．CISC はデスクトップパソコンや，スーパーコンピュータなどの動作速度が求められる環境で使用されています．

ARM は Advanced RISC Machines の頭文字を取ったもので，RISC アーキテクチャの一種でした．ARM は一般的な RISC アーキテクチャと比べ，低消費電力化に特化した設計になっており，特にモバイルデバイスのバッテリー寿命を延ばすための最適化が行われています．ARM はスマートフォンやノートパソコン，携帯ゲーム機などのモバイルデバイス，他にはデータロガーやセンサネットワークなどの IoT 機器にも使用されています．

私の研究テーマである「逐次比較型 AD 変換器の低消費電力化」との関連として，仮にシステム全体の構築を考えたとき，デジタルフィルタや，無線通信を実現するためには、CPU を搭載しなければならないこともあると考えられるので，使用するアーキテクチャとしては「低消費電力化」が利点となる ARM アーキテクチャを使用することになると考えられます．

私はハードウェア寄りのテーマを研究しているため，普段はソフトウェア側のことを考える機会があまりありませんでしたが，今回の課題を通して，実際に社会で役に立つシステムを構築するためには，多面的な視点をもち，様々な角度から物事を考えなければならないと感じました．また，釜江先生の経験談や日本の外の環境の話などはなかなか聞く機会のないお話だったので，とても興味深かったです．私は大学院に進学するため就職までに少し時間があるため，英語の勉強を進めて，グローバルに活躍できる人材になれるよう頑張ります．今回は貴重な機会を頂き，本当にありがとうございました．

報告書 3 電子情報工学科　中島良樹

私は宇宙のデータを Phython をつかって plot してみました．データはガイアアーカイブから範囲を決めて抽出して Phython で図示することができたので，当初の目的は達成しました．

今後は Phython をもっと使えるように勉強したいと思っています．現在，具体的な課題を見つけているところですが，Python はとても便利なツールだと，この活動を通じてわかったので，これを活用していろいろなことに挑戦していきたいと考えています．

報告書 4 電子情報工学科　中武大地

私が今回のテクノクラブに参加したきっかけは，テクノクラブの紹介で宇宙をガンマ線で観測した画像を見て，もともと宇宙に興味があったのでそれに関する何かを経験し

てみたいと思ったからです．最初に釜江先生の資料を拝見して，英語であったり専門的な内容であったり，自分の力不足を痛感しました。そこで同じテクノクラブの先輩や仲間にアドバイスなどをいただき，ガイアアーカイブやパイソンの使い方を少しずつ学びました．ただ調べるだけでなくやってみないと何も進まないということを，この経験や夏休み前の釜江先生とのズームミーティングで学びました。そのミーティングでは，釜江先生と鈴木先生が視野を広くとにかくなんでもやってみなさいと仰っていたのが心に残っています。

今回，私は最終的にガイアアーカイブから fits ファイルをダウンロードして，パイソンで M46という散開星団をプロットしました．ガイアアーカイブから必要なデータを絞ってダウンロードしたり，パイソンでの基本的なデータの読み込み方や座標の変換，for 文でのデータの一括プロットを主に学びました.

このような機会を得て，今まで触れてこなかったことに取り組めて，とても貴重なものになりました．発表のミーティングでは先輩の研究について少しだけ目に触れることもできて楽しかったです．また，パイソンは音声変換などで学びのサポートができることがわかりました.