Asia-Pacific Economic Cooperation

Advancing Free Trade for Asia-Pacific Prosperity

Developing Computational Thinking on AI and Big Data Era for Digital Society

Recommendations from APEC InMside I Project

APEC Human Resources Development Working Group

March 2021

APEC Project: HRD 01 2018

Produced by

Masami Isoda (Japan), Prof/PhD, University of Tsukuba

TEL: +81-298537283

Email: isoda@criced.tsukuba.ac.jp

Roberto Araya (Chile), Prof/PhD, University of Chile

Email: roberto.araya.schulz@gmail.com

Maitree Inprasitha (Thailand), Vice President/PhD, Khon Kaen University

Email: imaitr@kku.ac.th

For

Asia-Pacific Economic Cooperation Secretariat

35 Heng Mui Keng Terrace

Singapore 119616

Tel: (65) 68919 600

Fax: (65) 68919 690

Email: info@apec.org

Website: www.apec.org

APEC#221-HR-01.1

Preface

This is the recommendation towards educational reform for APEC economies from APEC Project HRD 01 2018 'Inclusive Mathematics for Sustainability in a Digital Economy (InMside)' which was implemented under the APEC central fund and self-fund of Japan, Chile and Thailand. The project had the following components:

- Steering Meeting at Tokyo, 7-9 February 2019 by Ministry of Education, Culture, Sports, Science and Technology, Japan and University of Tsukuba
  http://www.criced.tsukuba.ac.jp/math/apec/apec2019/
- InMside Seminar at Viña del Mar, 2-4 May 2019 by Ministry of Education, Chile and Universidad de Chile with support of Pontificia Universidad Católica de Valparaíso
  http://www.criced.tsukuba.ac.jp/math/apec/apec2019chile/
- Synthesis Meeting at Khon Kaen, 18-20 November 2019 by Ministry of Education, Thailand and Khon Kaen University
  https://www.crme.kku.ac.th/detail_page/apec2019.html

The recommendations were finalized by the Project Overseers and key actors with the contributions and collaborations of curriculum specialists from APEC economies on these seminar and meetings. The InMside has consecutive projects from secondary level (HRD 01 2018) to middle (HRD 05 2019) and primary education levels. Thus, InMside project HRD 01 2018 is called InMside I in this report. This recommendation illustrates the general framework for InMside projects by focusing on secondary school level curriculum reform for InMside I, HRD 01 2018.

Project Overseers of APEC HRD 01 2018
Masami Isoda, Prof/PhD, University of Tsukuba
Roberto Araya, Prof/PhD, University of Chile
Maitree Inprasitha, Vice President/PhD, Khon Kaen University
With support of Sofian TajulArus and Orlando Gonzalez

Acknowledgement

Project Overseers would like to acknowledge the support to the projects by the APEC Secretariat, Ministries of Japan, Chile and Thailand as well as curriculum specialists who deeply contributed at the seminar and meetings. All contributors are appeared on the following website with their contributions:

http://www.criced.tsukuba.ac.jp/math/apec/apec2019/

http://www.criced.tsukuba.ac.jp/math/apec/apec2019chile/

https://www.crme.kku.ac.th/detail_page/apec2019.html

TABLE OF CONTENTS

Chapter 1

General Issues for Curriculum Development

1. Introduction

In recent years, under the 4th industrial revolution by using AI and Big Data, APEC has been promoting the movement for regional reform in education. In 2017, APEC leaders pledged to work together to recognize the potential of the internet and digital economy and welcomed the adoption of the APEC Internet and Digital Economy Roadmap. On the Action Plan of the APEC Education Strategy for 2030 in HRDWG (2017), curriculum reforms are a key for the long-term development of human capital that ensures their mobility.

The project 'Inclusive Mathematics for Sustainability in a Digital Economy (InMside)' was proposed by Japan, Chile and Thailand with co-sponsoring economies: Indonesia, Korea, Papua New Guinea, Russia, Chinese Taipei, Malaysia, Viet Nam, China and Canada. It was approved under the APEC GPA funding priorities for 2018: "Promoting Sustainable, Innovative and Inclusive Growth", "Improving Connectivity", and "Innovation in the Digital Age", and treats model-tasks for "Enhancing Food Security and Sustainable Agriculture in Response to Climate Change", and "Strengthening Inclusive Growth through Structural Reform". Additionally, the project was done under APEC Chile 2019, with priorities focused on Digital Society, Integration 4.0, Inclusive Growth and Sustainable Growth.

Seeking the essential competencies for Digital Society in APEC region by using AI and Big Data, the project InMside focused on Computational Thinking in relation to using AI and coding, and Statistical Thinking in relation to using Big Data. On academic subjects, Computational Thinking is mostly related with computer science and Statistical Thinking is mostly related with statistics and probability. However the way of thinking is not limited to be used in specific academic subjects such as informatics at senior secondary school level but used in daily life. Thus, this report is not focused on curriculum sequence of specific academic subjects but discusses Computational Thinking and Statistical Thinking as general issues in later chapters. The major question of this introductory chapter is why InMside I project focus on Computational and Statistical Thinking for this curriculum reform, instead of just coding or programing thinking. For this question, current curriculum reform status and trends are discussed in this chapter.

2. General Curriculum Framework for 21st Century Skills

20 years ago, on the early 21st century, increasing the global economy by using various international networks enhanced by the Internet changed the world. At that time of globalization, the nature of the society became into a more competitive and challenging one. On this changing society, OECD (2005) proposed the competency for successful life and well-functioning society. Currently, every economy

engages on curriculum implementation for 21st Century Skills. At the HRD-Ednet Meeting in Chile (2019), Dr. Habibah Binti Avdul Rahaim, from Malaysia, proposed the following framework (Figure 1). Here, the 21st Century Skills were explained by Cognitive Skills, Soft Skills and Hard Skills. It is very much related with other economies' curriculum frameworks. For example, Japan (2017) has three pillars: (1) Ensure that knowledge and skills are acquired (Hard skill); (2) Develop the pupils' abilities to think, make judgements and express themselves (Cognitive skill); (3) Cultivate the will to learn and humanity (Soft Skill). [1] (See, Kano, 2019[2]; Noshiro, 2019[3]; Kawasaki 2019[4])

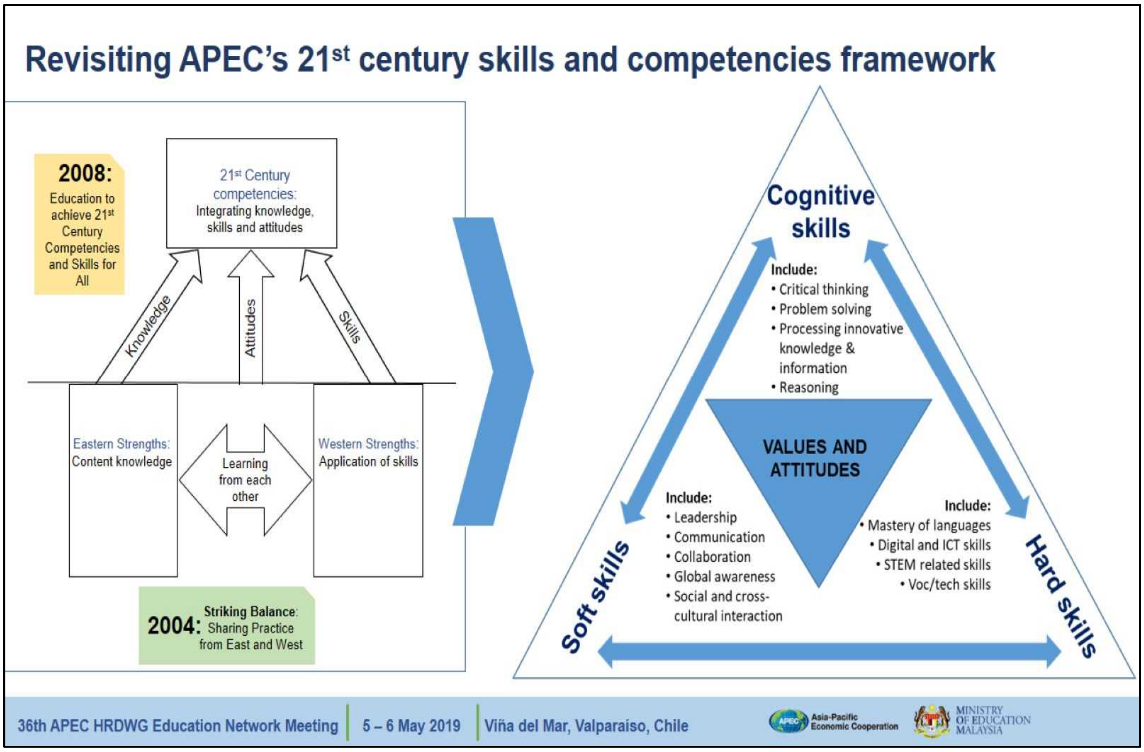


Figure 1. 21st Century Curriculum Framework by Dr. Habibah (2019)

These three pillars can be seen at the subject levels such as in the South East Asia Basic Education Standards for Mathematics and Science (the case for mathematics in Figure 2; SEA-BES, CCRLS, 2017) which is the curriculum document for Southeast Asia[5]. On Figure 2, soft skills are explained by Values, Attitudes and Habits for Human Character. In the 2000s, in relation to curriculum design, Bloom´s Taxonomy used to revisit for clarifying higher order thinking and there were critical discussions. On Figure 2, higher order thinking is clarified in mathematics education for the human character formation as the general aim of education.

These current frameworks were general frameworks as a response to the OECD competency but even though they work as general frameworks, they were previously established by focusing on the 4th

[1] https://www.mext.go.jp/content/20200227-mxt_kyoiku02-100002604_1.pdf
[2] http://www.criced.tsukuba.ac.jp/math/apec/apec2019/presentations/7Feb/2/rev2-Toshiharu_Kano20190207.pdf
[3] http://www.criced.tsukuba.ac.jp/math/apec/apec2019chile/
[4] http://www.criced.tsukuba.ac.jp/math/apec/apec2019chile/pdf/20190504-01.pdf
[5] http://www.recsam.edu.my/sub_SEA-BES/images/docs/CCRLSReport.pdf

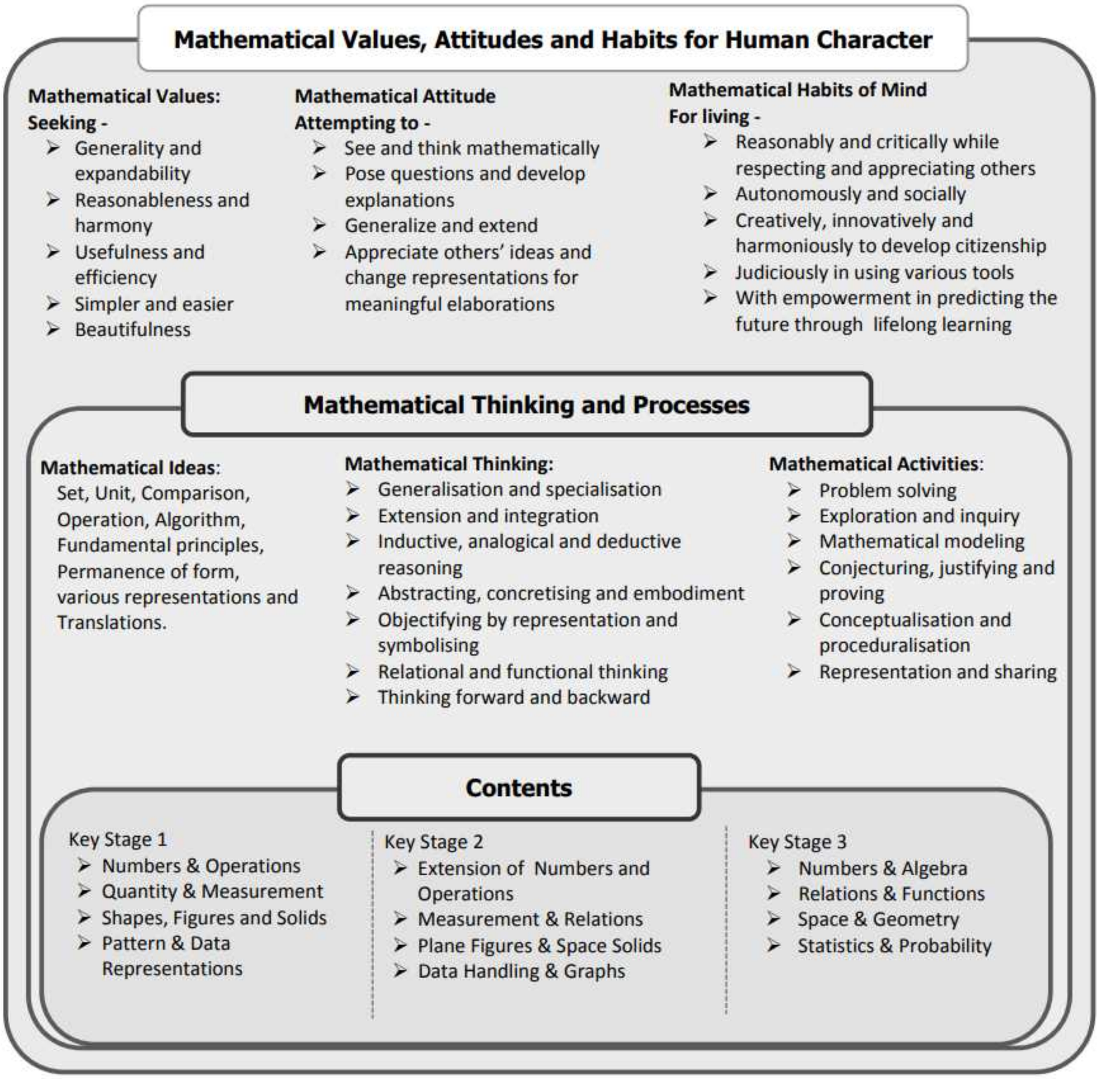


Figure 2. SEA-BES, CCRLS for mathematics (2017)

<table>
<tr><th>Logical Thinking</th><th>Critical Thinking</th><th>Design Thinking</th></tr>
<tr><td colspan="2">Logic</td><td>Intuition</td></tr>
<tr><td colspan="2">Persuasion</td><td>Huristics</td></tr>
<tr><td colspan="2">Mainly right side brain</td><td>Brain and Five senses</td></tr>
<tr><td>Check the logical truth</td><td>Subjective to Objective, Various perspectives</td><td>Leaps by Idea driven, hypothetical</td></tr>
</table>

Figure 3. Design Thinking VS Logical and Critical Thinking by Young and Kojima

Industrial Revolution under the progress of AI and Big Data.

3. Towards the Curriculum Framework for AI and Big Data era: Necessity of Humanity

On the 2010s, there were two related movements in education for AI and Big Data. The first movement was the Internet of Things (IoT) which is related with the design revolutions from Science, Technology, Engineering and Mathematics (STEM) Education to STEAM (A: Art) education movement. The second movement was the 4^{th} Industrial Revolution by AI and Big Data.

The IoT movement enhanced design as well as computer and technology. Based on Young and Kijima (2019), there was a movement from STEM education to STEAM education in the early 2010s which was led by John Maeda (2011). He enhanced the creation by simplicity and fostered the community of designers 'Aesthetics + Computation' (Maeda, 2004), and supported the Scratch language project at MIT which is currently used around the world. In the 2000s, there were problems of commonality on the Master of Business Administration (MBA) Program and difficulty for innovative busines creation. As a result, he produced a new MBA program with design thinking and collaborated with members of USA-Diet. Young and Kojima summarized the difference of the design thinking with other thinking which was enhanced at the new MBA program in Figure 3. A requirement to enter the MBA program was the GMAT (Graduate Management Admission Test) which asks one quick selection answer for English and Mathematical tasks to check the simple-quick decision-making skills to become an MBA program student. On the basis of English and Mathematical skills, they learned how to produce and manage innovative business for the 4^{th} Industrial Revolution under the AI and Big Data era.

Educational Reform on the 4^{th} Industrial Revolution is summarized by UNESCO (2019) as the Beijing Consensus[6], which recommends governments and other stakeholders in UNESCO's Member States to:

- Plan AI in education policies in response to the opportunities and challenges AI technologies bring, from a whole-government, multi-stakeholder, and inter-sectoral approach, that also allow for setting up local strategic priorities to achieve SDG 4 targets.
- Support the development of new models enabled by AI technologies for delivering education and training where the benefits clearly outweigh the risks,

[6] https://en.unesco.org/news/first-ever-consensus-artificial-intelligence-and-education-published-unesco

and use AI tools to offer lifelong learning systems which enable personalized learning anytime, anywhere, for anyone.

- Consider the use of relevant data where appropriate to drive the development of evidence-based policy planning.
- Ensure AI technologies are used to empower teachers rather than replace them, and develop appropriate capacity-building programmes for teachers to work alongside AI systems.
- Prepare the next generation of existing workforce with the values and skills for life and work most relevant in the AI era.
- Promote equitable and inclusive use of AI irrespective of disability, social or economic status, ethnic or cultural background or geographical location, with a strong emphasis on gender equality, as well as ensure ethical, transparent and auditable uses of educational data.

On APEC InMside I project, the major point of discussion for the new curriculum framework is focused on humanity and coding, instead of just coding. Following the UNESCO Beijing Consensus, the objective is to prepare the next generation of existing workforce with the values and skills for life and work most relevant in the AI era. Soft and cognitive skills are necessary even through people are required to have a hard skill for coding.

Isoda (2018) proposed three types of Humanistic-Social Competency for using computational thinking and statistical thinking as follows[7]:

a. **Competency to lead** through creating the new value on the infrastructural bases for Society 5.0[8], beyond the 4th Industrial Revolution.
b. **Competency to work** through predicting the future society with negative and positive aspects under the 4th Industrial Revolution and SDGs, and utilize AI and Robots judiciously and critically to develop well-functioning society.
c. **Competency to live** finding the own value for humanistic oriented activity on the Digital Society. AI and Robots are progressive technology and tentatively functioning on their limitation even when we don’t recognize it as a part of our life. It creates the new progressive-humanistic roles for us.

The report of the Ministry of Economy, Trade and Industry (METI), Japan (2019) states Top human capital training as (a), Middle human capital transfer as (b), and IT literacy improver as (c).[9] For (a), METI reported the Era of Mathematical Capitalism which

7 https://kaken.nii.ac.jp/ja/grant/KAKENHI-PROJECT-19K21743/
8 https://www.japan.go.jp/abenomics/_userdata/abenomics/pdf/society_5.0.pdf
9 https://www.meti.go.jp/shingikai/economy/jinzai_ikusei/pdf/001_03_00.pdf

enhances mathematical science and includes Informatics as the most necessary science for economic development. It enhanced the necessity of high qualified mathematics education for (a).[10] This is related with the documents which enhanced the necessity of mathematics such as the Era of Mathematics from the Engineering and Physical Sciences Research Council (2015) in UK.[11] The Australian Academy of Science (2015) also published 'The importance of advanced physical and mathematical sciences to the Australian economy.'[12] Those documents[13] enhance the competency of mathematics in relation to Informatics and Mathematical Sciences such as Data Analysis, Modeling and Simulation for using AI and Big Data.

Lew (2020) analyzed those three human capitals from the functionality of AI.[14]

- Humanistic perspective: AI is a human heritage. Students need to understand the key skills that guide the rapidly change of the modern period.
- Ordinary life perspective: AI is closely related to our daily lives. Students need to learn AI in order to get on with their daily lives.
- Developmental perspective: AI is like a problem solver. Students can develop their own thinking power by exploring basic knowledge to operate AI.
- Social Developmental Perspective: AI is changing the current society as well as will change our future society. Students have to learn the AI technology to speed up the development of workstation where they will contribute as workers after graduation from school.

Thipakorn (2019) summarized the discussion of the seminar in Chile on InMside I by the components on Figure 4.

TajulArus (2019) summarized the discussion of the Synthesis Meeting in Thailand by Figure 5 which includes both computational thinking and statistical thinking for InMside.

10 https://www.meti.go.jp/shingikai/economy/risukei_jinzai/pdf/20190326b_report.pdf
11 https://epsrc.ukri.org/newsevents/pubs/era-of-maths/
12 https://www.science.org.au/supporting-science/science-sector-analysis/reports-and-publications/importance-advanced-physical-and
13 Other documents: "MATHEMATICAL SCIENCES: DRIVING THE UK ECONOMY", Council for Mathematical Sciences, 2016; "One Step Beyond: Making the most of postgraduate education", Department for Business Innovation and Skills (BIS), 2010; "THE ERA OF MATHEMATICS". The Engineering and Physical Sciences Research Council (EPSRC), 2018; "A Study of the Socio-Economical impact of Mathematics in France", AMIES, 2015; "FOR A MEANINGFUL ARTIFICIAL INTELLIGENCE", Cedric Villani, 2018; "Mathematical sciences and their value for the Dutch economy", Platform Wiskunde Nederland, 2014; and "THE IMPORTANCE OF ADVANCED PHYSICAL AND MATHEMATICAL SCIENCES TO THE AUSTRALIAN ECONOMY", Australian Academy of Science, 2015
14 http://www.criced.tsukuba.ac.jp/math/apec/apec2020/

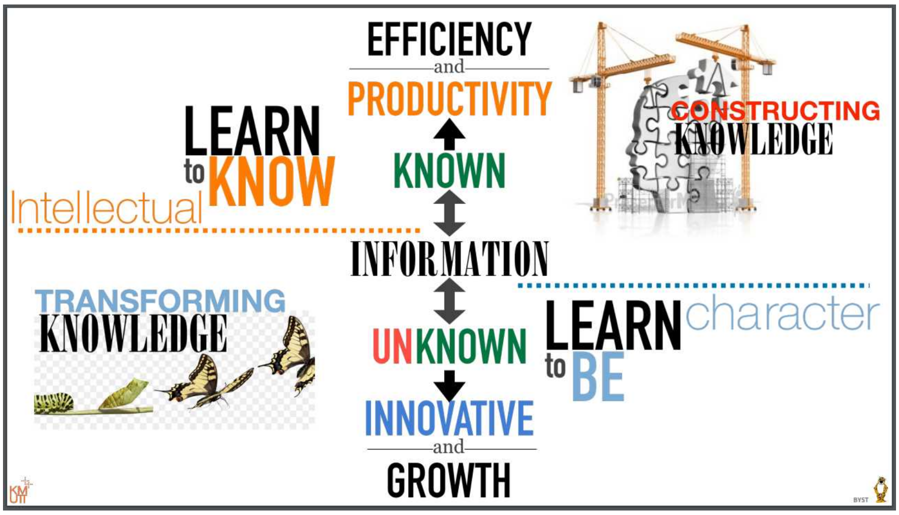


Figure 4. For Humanistic Components for Era of AI and Big Data (Thipakorn, 2019)

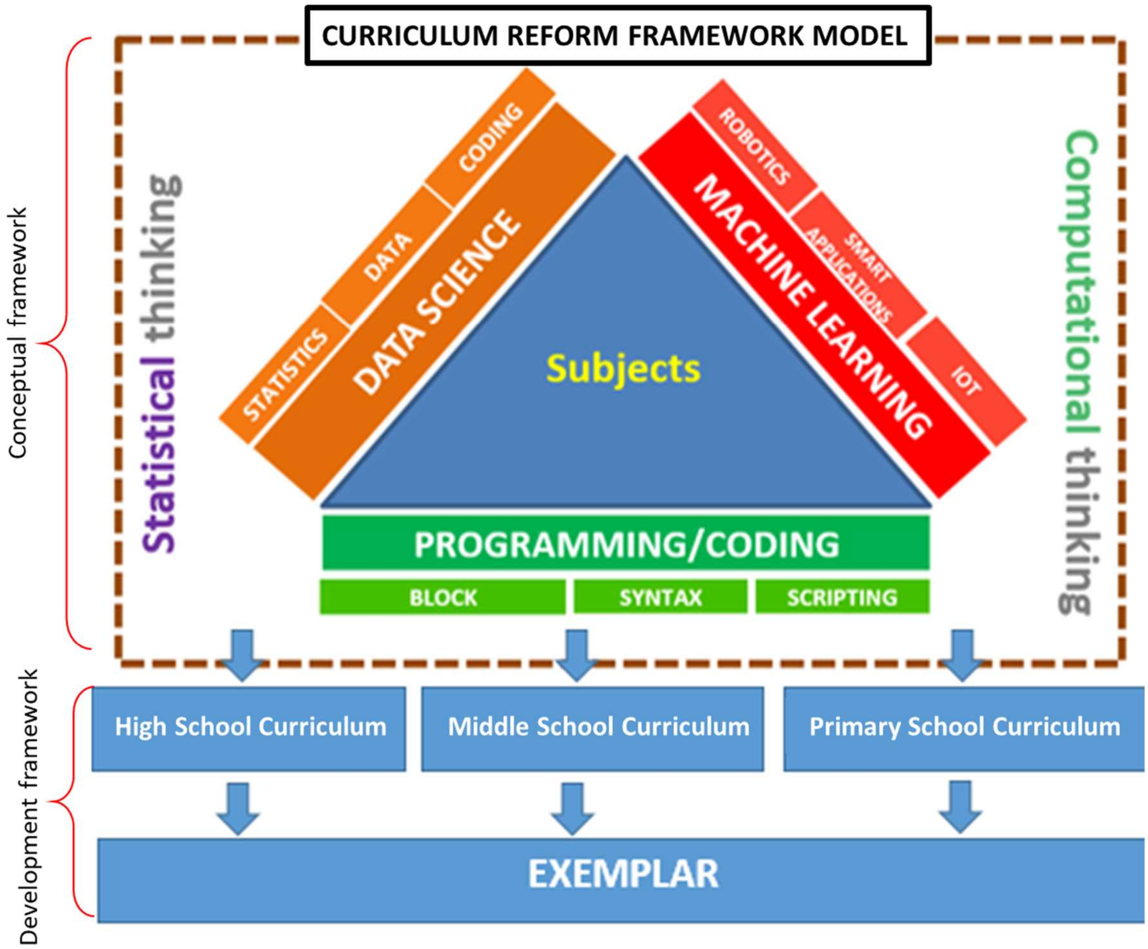


Figure 5. Curriculum Framework for Era of AI and Big Data on InMside

Figure 5 is the general framework for Computational Thinking and Statistical Thinking for InMside which implicates that any curriculum materials in education from K to 12 can be seen from both ways of thinking and propose the necessity to revise the curriculum on this perspective to develop Computational Thinking and Statistical Thinking through education. The revision of the framework for Computational Thinking and Statistical Thinking will be explained at the chapters 2 and 3. Even though it looks like humanity is not included, there is a reference to 21st century skills curriculum framework in general and specific such as Figure 1 and Figure 2.

(Wording of Chapter 1 by Masami Isoda)

Chapter 2.

Recommendations for Curriculum Reform on Computational Thinking

Roberto Araya — CIAE IE, University of Chile, Chile
Masami Isoda — University of Tsukuba, Japan

In this chapter we propose recommendations and exemplars for a curriculum on Computational Thinking for the Digital Economy (Isoda, 2020). In order to do so, we consider the challenges our students face on the 21st century for the new coming digital society. We consider a vision proposed by Thipakorn (2020) based on three central questions: What characters do our students need to "be"? What competences do our students need to be able to "do"? What literacies do our students need to "know"?

1. INTRODUCTION

Already smartphones efficiently guide us to reach a place in the city and will soon take control of our vehicles, negotiate with other devices, decide for us to undergo a surgical operation, and they will choose our career and friends. We are just entering the digital society and it is still difficult to foresee several other disruptive transformations that will change the nature of our personal and professional lives. What characters do our students need to "be" on the digital society? Is Computational Thought a new type of thinking that we did not use until now and why should we help develop in new generations?

Why Computational Thinking (CT)?

World organizations such as UNESCO, countries, and scholars are increasingly proposing that Computational Thinking (CT) is a critical cognitive disposition that students need to adopt and use it efficiently to solve the challenges of the digital society. For example, Conrad Wolfram recommends that "we should be teaching math as if computers exist." We can add that we should also teach biology as if computers have enormous power in genetic analysis. We should teach social sciences considering the availability of an enormous flow of people's behavior data coming from their mobile phones. We should teach language considering the disruptive advance of machine translation and voice recognition. Yet, we teach all subject matters as if this enormous and constantly increasing computational power did not exist! Most problems today are solved using computers, from servers on the cloud to our smartphones. CT is necessary to have the ability to tap on computers and modern digital tools to work to solve those problems. Koshi (2020) emphasizes that in the digital society it is necessary to have the qualities and abilities to observe and assimilate various phenomena in the world as information and connections and utilize information and information technology appropriately and effectively to discover and solve problems and form their own ideas. Therefore CT must play a very important role in the new curriculum.

What is Computational Thinking?

CT comes from an aspiration that has been turning for centuries and it is beautifully represented in a letter written by Leibnitz in 1685 “The only way to rectify our reasoning is to make them as tangible as those of the mathematicians, so that we can find our error at a glance, and when there are disputes among persons, we can simply say: Let us calculate, without further ado, to see who is right”. Leibnitz was influenced by the work of 14th century Majorcan philosopher, Ramon Llull, who designed a mechanical machine to reformulate arguments and ideas in terms of a characteristica universalis, or universal language, so as to have computable arguments. According to Llull, the machine could prove for itself the truth or the lie of a postulate. This means, decomposing arguments in term of thousands of simple units, which can be recombined and thus able to be expressed and performed as mechanical computations. Later, in 1936 Alan Turing proposed a sequential machine, the Turing machine, which provides a precise definition for computational steps and algorithms. This is the core of computational thinking as expressed today by computer scientist Jeannette Wing (2006): “Computational Thinking is the thought processes involved in formulating a problem and expressing its solution(s) in such a way that a computer – human or machine – can effectively carry out”. In other words, CT is understood as a problem solving paradigm of mathematizing a problem in such a way that the computer can execute it (Tin Lam, 2020). CT moves students beyond technology literacy, it presents endless possibilities for creatively solving problems (Nguyen Chi Thanh). Think, for example, the Tower of Hanoi puzzle, proposed by the French mathematician Edouard Lucas in 1883. It is a problem already classic and widely used in computer science courses to begin to develop algorithmic thinking, and very particularly to introduce the notion of recursion.

Is this reasoning different from mathematical thinking (Isoda et al, 2012)? CT is very powerful for problem-solving. It includes several elements (Ulep, 2020). First, it requires being able to formulate problems in a way that enables to use a computer and other tools to help solve them. Second, the student must be able to represent data logically. This means, ordering, organizing and analyzing data. Third, it is also needed to be able to select and preprocess the data. Fourth, students have to have the skills to build models and do simulations. Fifth, students have to develop the attitude to automate solutions, and do it through algorithmic thinking (a series of ordered steps). Sixth, they have to identify analyze, and implement the most efficient and effective solutions. Seventh, students have to be able to generalize and transfer this problem-solving process to a wide variety of problems. Programing Thinking cultivated through repetitions of these steps (Kano, 2020), and is illustrated in figure 2 adapted from Kano (2020).

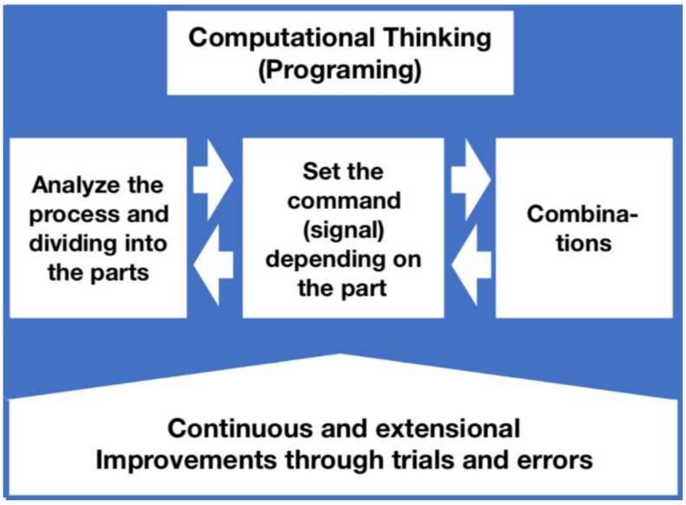


Figure 1: Programming Thinking

Four basic strategies of Computational are decomposition, pattern recognition, algorithms, and abstraction. Decomposition means divide to conquer (Jordan, 2013). The problem has to be considered as a sequence of simpler problems. These simpler problems have to simple enough so that can be solved mechanically, and hopefully in parallel. This is critical for having good scalability. Pattern Recognition means to be able to detect and describe powerful patterns that can contain the mechanisms to find a solution of the problem. Algorithm is the basis of algorithmic thinking. This occurs when someone observes repeated patterns in problems and then generalizes a set of rules for dealing with such situations (so that one need not think this through anew each time that problem occurs). Abstraction means to extract some hidden characteristics that are critical for the problem, and simplify them to be able to manage them without losing its power. Additionally, we have to consider that all these types of thinking are not operating isolated from each other, they have to be interconnected and successfully integrated.

What students need to be?

Computational thinking has close connections with epistemology, ethics and wisdom. All of them share several core concepts such as knowledge, learning, and values. For example, CT provide tools to design intelligent and moral artificial agents that accompanies and give emotional and intelligent support to us. These are agents that can be completely autonomous, but we have to learn to trust them. Consider the case of an accident in an autonomous vehicle. Who is responsible? What happens if the accident is due to a system error, or due to an error in design flaws? But what happens if the vehicle learns autonomously?

Another critical aspect is moral biases. Devices learn autonomously, but as us, they are learning from information that already has strong sexual and ethnic biases. Students have to be conscious and vigilant on this issues, understand why they could emerge, and able to do the corrections required. They have to have the ability to grasp various phenomena from the viewpoint of information and connections, and connect multiple pieces of information to find new meanings, and attitude toward creating information technology to improve quality of life and build a sustainable society (Koshi, 2020).

2. WHAT STUDETS NEED TO DO?

There are three core Computational Thinking processes students need to do (Araya et al, 2019): algorithmic thinking, also called traditional computational thinking (Santillan, 2020), computational modeling, and machine learning. Each one of them provides the basis for becoming an empowered citizen on the digital society. They are not sufficient, but nevertheless necessaries.

Algorithmic Thinking

Algorithmic thinking emerges when someone observes repeated patterns in problems and then generalizes a set of rules for dealing with such situations (so that one need not think this through anew each time that problem occurs). In effect, one creates (formally or informally) an algorithm or rational procedure. This sort of thinking is often used in problem solving and in computer programming. It is the most easily recognizable pillar of computational thinking.

According to Chi Thanh (2020), algorithmic thinking can be very well illustrated and promoted with the Tower of Hanoi game. In this game the student has to start on the configuration on the left of figure 2, and her goal is to obtain the configuration of the right. To do that she has to move beds from one mast to a neighbor one making sure that always smaller beds are above bigger ones.

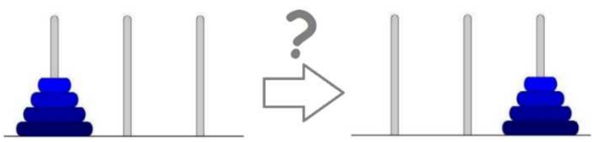


Figure 2: Tower of Hanoi game

The three initial moves will be the following

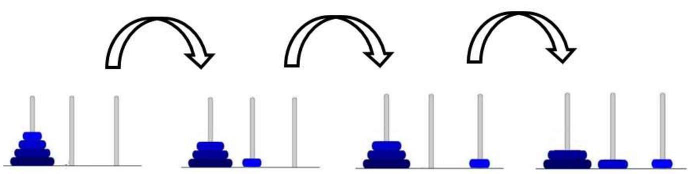


Figure 3: Three initial moves in the Tower of Hanoi game

These three moves, shown in figure 3, satisfy the game restrictions, but the student has to think how to continue and whether is possible to reach the goal.

What instructions on paper or in your smartphone you would write and send to a classmate in order to solve this challenge? What is the shortest set of instructions that you could write? Is there a certain pattern that you can capture for writing efficient instructions?

Computational modeling

A computational model is a mathematical model which implies and uses computational means. For example, the following sequence of boards represent a forest fire propagation model (Araya, 2017). Each cell is an acre that can be burning (red), already burned (black) or un-touched yet by the fire (white). The left board shown in figure 4, is at the moment of initial fire. The second board illustrates the result of the propagation to the next time period. The third board shows what the status at the next time period.

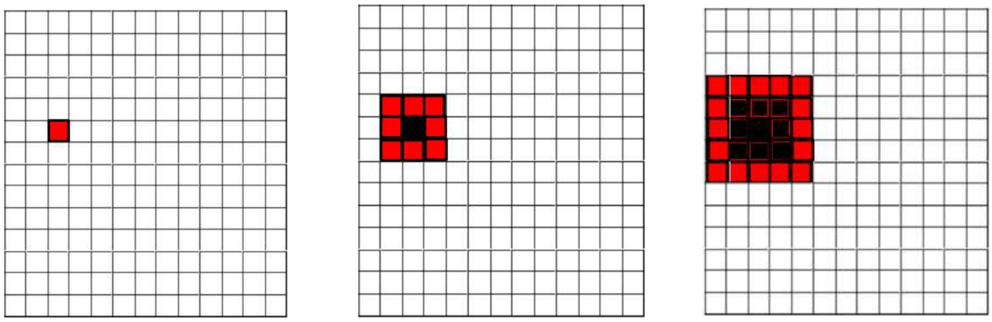


Figure 4: three consecutives stages of a forest fire propagation

The model specifies mechanical and computational rules that describe the mechanism of fire propagation.

For each cell of the board write the rules that specify its behavior. How should each cell be painted on each time period? How would you modify the rules to include wind direction at each cell?

Machine learning thinking

Machine learning (ML) is the science algorithms used to learn to perform a specific task without using explicit instructions, relying on set of examples, detecting patterns and doing inferences. It is a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training set", in order to make predictions or decisions without being explicitly programmed to perform the task. According to Jordan (2019), ML is an algorithmic field that blends ideas from statistics, computer science and many other disciplines to design algorithms that process data, make predictions, and help make decisions. Thus ML is also part of statistical thinking (Gonzalez et al, 2019), and requires probabilistic intuition and math knowledge (Visostskiy, 2020).

ML requires recognizing categories and the list of features that can potentially predict the categories, the recollection of examples or cases, the use of different machine learning algorithms, and the ability to test the validity and interpret the results. Machine Learning is at the core of Artificial Intelligence, because it is learning by machines and because of its applications to real world problems. UNESCO (2019), recently reviewed the recent trends in the evolution of AI and its profound impact on human societies, economies and the labor market, as well as on education and lifelong learning systems. UNESCO recommends governments to consider planning AI in education policies

Typically the data set is divided in two: training set and test set. Using the data from the training set the ML model are built. Then using the test set the models are run on that set and assessed the quality of its predictions.

One key component to understand ML algorithms is scatter plots as in figure 5, where two classes are displayed: circles and crosses. These form two clouds on the scatter graph.

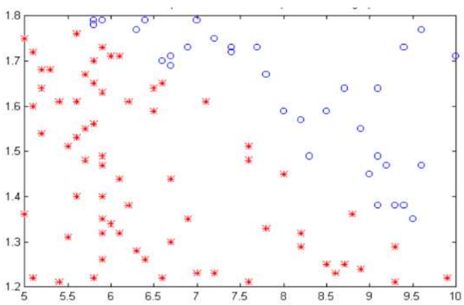


Figure 5: scatter plot with two classes: red crosses and blue circles

The core idea of ML is finding a curve that better discriminate the two clouds as shown in figure 6.

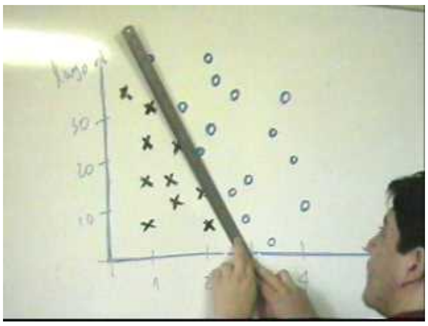


Figure 6: Student placing a ruler in order to better separate the two point clouds

In our previous example, a line discriminating both clouds is shown in figure 7.

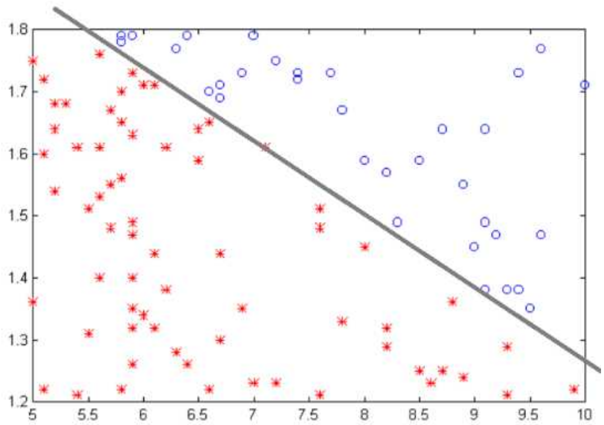


Figure 7: Straight line that best separates the two clouds

This straight line is a linear discriminator that can be obtained using Support Vector Machine (SVM) method. This is an example of machine learning a linear model. In a two dimensional scatter graph, this means finding a straight line that better separates the two clouds, as shown in figure 8 taken from the Open Lesson at the Chilean National Congress, May 2, with an 8th grade class, for the APEC 2019

InMside meeting in Chile.

ML is part of statistics and data science. It combines domain expertise, programming skills, and knowledge of math and statistics to extract meaningful insights from data, as for example in (Lambas, 2020) for the analysis of India air quality data. Data science practitioners apply machine learning algorithms to numbers, text, images, video, audio, and more to produce artificial intelligence (AI) systems that perform tasks which ordinarily require human intelligence. Students need to be exposed to data interpretation in which they are required to apply their theoretical knowledge to numbers and use their reasoning skills to make inferences based on logic and trends from given sets of data (TajulArus, 2020).

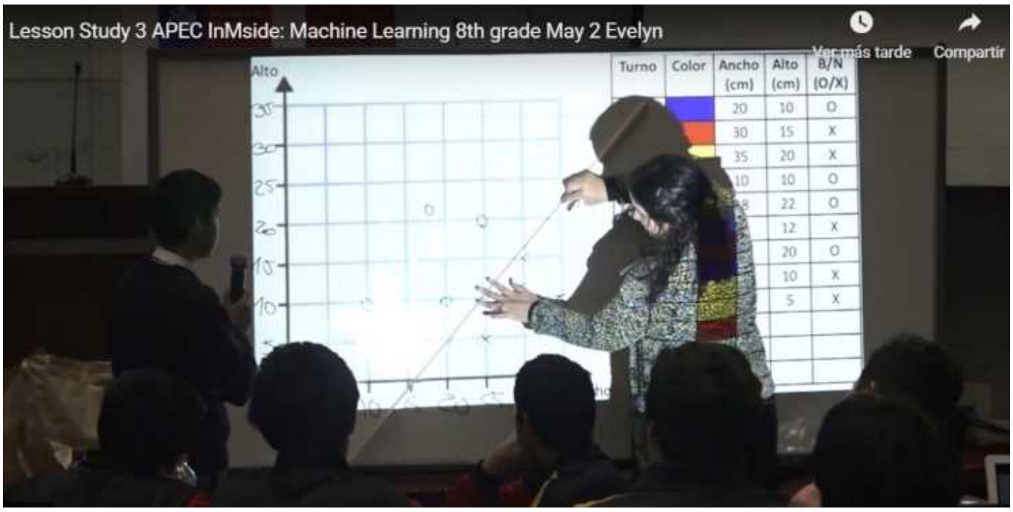


Figure 8: placing a ruler to best separate two clouds

For example, after detecting the most frequent characters on the Little Red Riding Hood story, the frequency table is

Granny	10
Mother	1
Little Red Riding Hood	5
Wolf	12
Woodcutter	2

Table 1: Frequency of characters named on the Little Red Riding Hood story

Moreover if we connect those characters that are in the same line in the book, the graph in figure 9 is obtained:

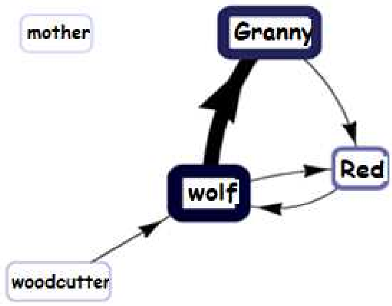


Figure 9: Graph obtained by co-occurrence of characters at the same line of the text.
This mechanically built graph summarizes the interactions of the story.

3. WHAT STUDENTS NEED TO KNOW?

As emphasized by Yunianto (2020) Computational thinking was conceived by Pappert as a thinking skill and "objects-to-think-with" that arises from working with computers. Here we consider the different strategies when doing CT using computers.

How to choose a computational thinking strategy?

When solving a problem we have to consider and assess first the amount of data available. If there is not much data then we have to consider the type of knowledge required to solve it. If it is mainly heuristics, then we have to specify the rules and build the code. For example, building expert systems with hundreds of rules. However if there are mathematical model that can be translated onto computer models then this type of models should be included. For example, models that represent dynamics with discrete time and discrete space. On the other hand, if there is huge amount of data, then a machine learning model can be built. In this case the system is not programmed but trained with the data. These decisions can be represented as a decision tree like the one in figure 10 (Reynolds and Araya, 1995).

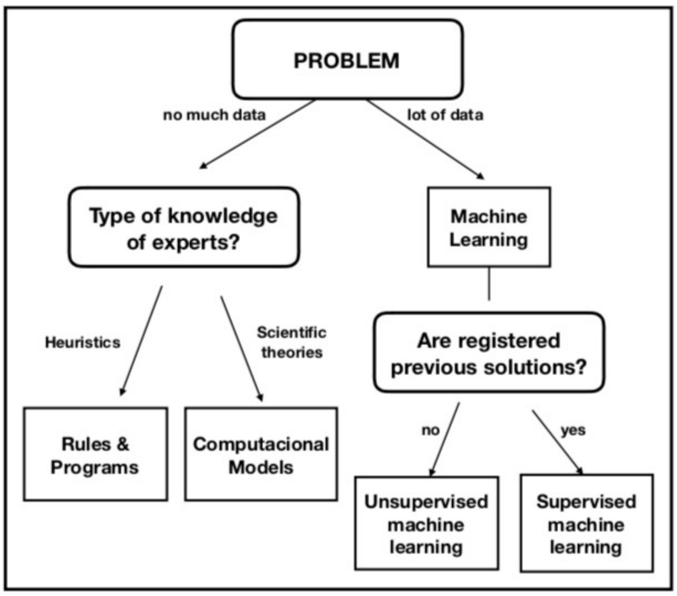


Figure 10: CT strategy selection decision tree

What is Artificial Intelligence?

Artificial intelligence (AI) is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans. Leading AI textbooks define the field as the study of "intelligent agents": any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals. AI usually combines Algorithmic Thinking. Computational Models and Machine Learning.

For example, the board of figure 11 illustrates the trajectory of an agent that moves maximizing the numbers on the board (Araya, submitted). This is the case of tropism of a simple animal moving to where is more food. In the following figure the agent starts in the blue cell with 10 units, then move to the neighbor cell with 15 units, then to the neighbor cell with 24 units, and so on.

The mechanism is very simple, but for an observer that doesn´t see the number of the cells the behavior can be interpreted as intelligent. Moreover, the numbers can be other values that represent odor and the agent is a dog searching for a hidden bone. The numbers can also be more abstract values (like degrees of risks) for an autonomous vehicle. A more challenging situation is when the numbers of the board change as the agent move. This means, the agent modifies the environment with its behavior. This is a more real model for a bigger number of applications. Additionally, if we have other agents, they also modify the environment. This scenario is much closer to the real world. Decision in such a case are very complex, and requires more sophisticated intelligence. Furthermore, if we add a more complex inner life to each agent then we approach to the behavior of organisms with higher cognitive and emotional mechanisms. Finally we get to agent that model human behavior, to whom we can talk and work together. Thus, AI and Humanity, start to be deeply connected. Both are necessary to develop a new literacy (Isoda, 2020).

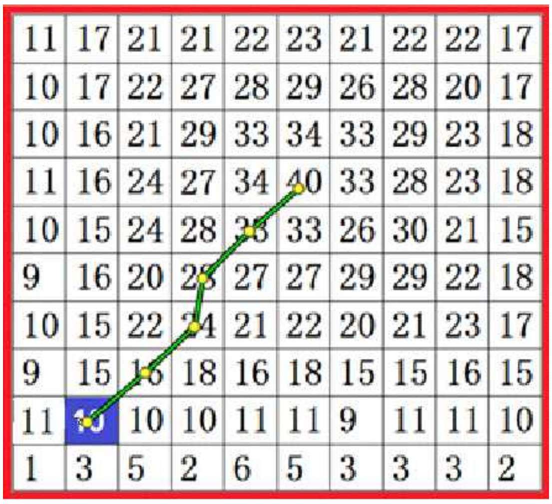


Figure 11: Board with numbers that represents nutrient units and organism trajectory

optimizing its nutrition

However there are still several limitations in AI that our students should be aware of. Kenyon (2019) emphasizes the inability of AI to learn by analogy. Analogical thinking is a critical ability in humans and several other animals. Another limitation are biases and the concentration in superficial features. This limitation is illustrated by Kenyon, in a visual categorization task to learn discriminate dogs from cats. If all dogs in the training set wear a metallic tag and cats do not wear such tags, then ML will learn to categorize based on the presence of the tags and not real differences between dogs and cats. This is an error that toddlers neither other animals do.

Why machine learning is so popular currently?

Machine Learning is more powerful with more data. Given the increasing availability of data, from the web, from instruments, cameras, smartphones, internet of things, transactions data from retail, financial and medical data, more and more applications are possible, and better precision is reached. On the other hand, the huge increase of computer power makes possible to handle more data and faster. This is critical for accurate learning.

According to Moore´s law; the number of transistors (computing power) duplicates every 2 years. This law has been valid since 1975 when Gordon Moore postulated. Additionally, there has been big discoveries on Machine Learning algorithms. These increasingly effective machine learning algorithms has already had a big impact on visual object recognition and on speech recognition that makes possible today automatic face recognition and automatic transcription from speech to text with human level error rates.

What Machine Learning methods students should learn?

There are several types of ML algorithms. The main ones are the following

Artificial neural networks

Artificial neural networks (ANNs), or connectionist systems, are computing systems inspired by the biological neural networks. Such systems of networks of abstract neurons are trained with examples or cases. Thus the network "learns" from this examples. This is done without being programmed with any task-specific rules.

Figure 12 shows a layer with two abstract neurons on the left and their connection to a third neuron on the right.

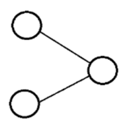


Figure 12: Two layers neural network

This is a very simple neural network, with three neurons and two connections. The activation of the left neurons produce an activation on the right neuron, and this depends on the strength of the connections. In the following case, shown in figure 13, one connection has more strength

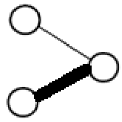


Figure 13: Two layers neural network with one connection with higher strength

Machine learning with neural networks is the process of changing the strength of the connections so that the network does a specific recognition or action task.

Figure 14 is a neural network with 6 neurons. At the left is the input layer with 2 sensory neurons, at the middle layer there are 3 neurons (this is called a hidden layer), and at the right there is a layer with 1 output neuron.

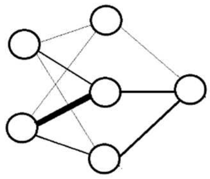


Figure 14: Three layers neural network

Decision trees

Decision tree learning uses a decision tree as a predictive model to go from features that are represented in the branches to predictions, which are represented in the leaves.

For example in figure 15 a decision tree the model predicts if the engine oil has water contamination or not. The feature is the concentration of Sodium (Na). If it is lower than 17 ppm, then the prediction is that the oil is normal. If the Na concentration is higher than 17 ppm then the prediction is that there is water contamination.

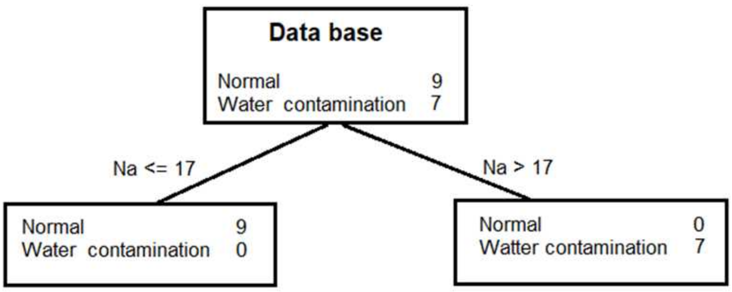


Figure 15: Engine diagnosis decision tree

Support vector machines (SVM)

An SVM training algorithm is a linear classifier as shown in figure 16. This means that the predictor is a linear combination of a sample of the features. For example, a prediction to be sick Z is a combination of fever temperature X and degree of pain Y. That is

$$Z = a*X + b*Y + c$$

The SVM algorithm finds appropriate coefficient *a, b* and *c*.

Figure 16: Linear classifier marked with a ruler

Genetic algorithms

A genetic algorithm (GA) is an algorithm that mimics the biological process of natural selection, where phenotypes are the solutions that depend on genotypes. After mutation and crossover new genotypes and therefore new phenotypes appear on the population, and some of these phenotypes can solve the problem.

For example, in strategy discovery, both conscious and unconscious, the process can be explained throughout a genetic algorithm mechanism (Siegler et al, 2005). In order to solve 25 + 12 – 12 one basic strategy is to focus attention at the left of the expression and then do 25+12. Then shift attention to the right and subtract 12. These actions can be expressed as composition of operators. In graphical terms this is a sequence of blocks (like Lego blocks) illustrated in figure 17. Each block is an action or a micro machine. Following the math convention for composition of operators, the first to be executed from right to left.

Figure 17: Sequence of actions executed from right to left that represent the strategy to first add the two numbers at the left of the expression and then add to that the number at the right

This sequence is like a molecule composed of genes. The yellow gene or action is O_Left that instructs the robot to focus visual attention on the extreme left of the expression. This means to focus on 25 on the expression 25+12-12. Then O_sRight instructs the robot to shift one position to the right. This means where the first 12 is. Then O_Sum instructs the robot to add the two numbers. That is, 25+12. Then O_sRight instructs the robot to shift the visual attention and focus on the second 12. That is -12. And finally the last action instructs the robot to add the previous result with -12.

How then is the genetic mechanism to discover an alternative strategy to solve this problem? If after a deletion and mutation, the insertion shown in figure 18 is made, then the sequence can be subject to the following transformation:

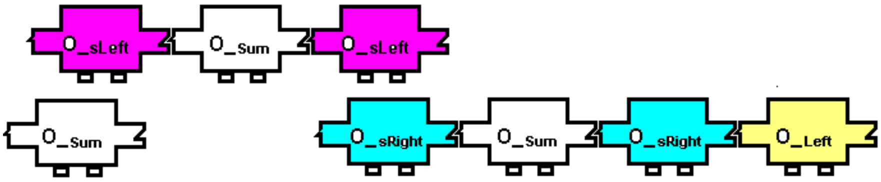


Figure 18: Insertion of the upper sequence in the original one

But this long sequence is redundant. It computes certain number that doesn´t affect the final result. If the redundancy is eliminated the sequence in figure 19 is obtained:

Figure 19: New sequence obtained after elimination of redundancy

But this sequence means go first to the 12 – 12 section and do the operation and then add 25. This is a new strategy, must efficient, and completely discovered by the genetic algorithm. This case illustrates how a new sequence of operations is discovered by mutations and crossover.

Another fruitful example is in design and architecture (DeLanda, 2012). For example, designing a song or a poster for a given market niche. Initially ten proposals are made by different designers. This is the first generation. Then the proposals are assessed by a random set of 200 people from the market niche in a web score system. Using this scoring, the 5 best are selected and extra 5 are generated mixing the best with the other 5 better proposals. This is the second generation of 10 proposals. Then another sample of 200 people score the proposals of the second generation, and the process continues for 8 generations.

Before the implementation of sophisticated learning models, it is good strategy to try mechanical versions of it (Wylie, 2018). For example, for a learning algorithm to play toc-tac-toe, Donald Michie implemented it with a large pile of matchboxes. An explanation and simulator of the mechanical game can be seen in http://www.mscroggs.co.uk/menace/ .

We suggest that to learn an algorithm it is a good didactic strategy to ask students to try to implement it in a concrete and mechanical way before writing code in a particular programing language.

4. CURRICULUM FRAMEWORK

Curriculum framework for traditional coding already exists, however it is needed a new one with the focus on core concepts of AI, computational modeling, and Machine Learning, and some of its typical applications.

Machine learning models are feasible to teach at Middle and high school?

Students can easily learn the idea of features and categories or predictions. Exemplar 2 illustrates these

concepts. We need students to learn to recognize in different applications what would be features and what are classes. For example in face recognition, features are just the color of each pixel of a photo, and the class is 1 if this picture is the picture of a given person and 0 if not.
Notion of training set and test set. Both sets are part of all the recollected cases. The students has to randomly select a proportion of the pictures, for example 50%, for training with ML. And once generated the algorithm, test accuracy on the rest of the examples.
Decision trees and SVM core ideas can be implemented in middle and high school using

- 1 dimensional scatter plots with the number line
- 2 dimensional scatter plots and
- Drawing discriminators with a rule
- Finding the equation of the linear discriminator
- Testing the discriminator in new testing set of data.

Curriculum framework

Tin Lam (2020) asks whether Computational Thinking should be a subject taught in school or to be incorporated in an existing subject. Ulap (2020), proposes that the goal of the K to 12 mathematics Curriculum is the development of learners' problem solving and critical thinking skills.
In order to materialize the new curriculum underlying the above said entities, a Curriculum Reform Framework model has to be established for the curriculum development process. This model will become the reference implementation framework model in carrying out the design process of curriculum formulation and implementation of the new curriculum proposed in each APEC member economies for high school, middle school and primary school respectively.
The proposed curriculum reform framework is based on the following implementation framework model (TajulArus, 2020).
The framework is divided into two main layers which are the conceptual framework and development framework (adapted from TajulArus, 2020), shown in figure 20.

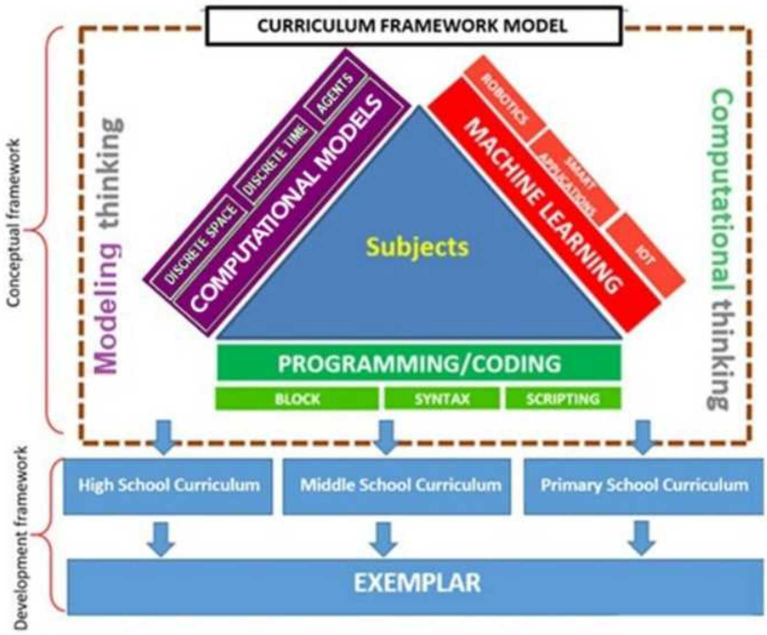


Figure 20: Curriculum Framework for Computational Thinking on InMside

The conceptual framework

It defines the important aspects to be addressed in the new curriculum based on the key elements underlined whether it becomes a subject-based curriculum, thematic, interdisciplinary or embedded across curricula. While it is also important for economies to decide on whether to adopt Standard-based Curriculum or Outcome-based Curriculum or Competency-based Curriculum.

The development framework

Where curriculum developers refer the implementation framework model and take the necessary actions in designing and developing the new curriculum at high school, middle school and primary school based on the following fundamental curriculum components:

Content

Refers to the syllabus developed on the content of the knowledge discipline which related with academic subject. On syllabus, the content is existed on the sequence and the scope of the content based on the learner's cognitive level. The respective knowledge, skills and values are determined accordingly.

Objective

Objective specified under general aim of curriculum and content on the syllabus.

Teaching and Learning materials

Teaching materials is the content which embedded the objective in curriculum. It includes tools such as hardware and software that are used to assist learners to meet the expectations for learning based on the specified curriculum.

Exemplars

For teacher, it is the main reference for curriculum development. It shows content and teaching materials in realizing the new curriculum through documented teaching and learning activities which covers the curriculum components. For students, it is the learning materials which are able to learn content and objective, and transfer various ways.

Methodology and Didactics

Refers to how knowledge or content is imparted or transferred to the learner. Methodology and didactics are chosen to realize authentic activity for students to produce and learn knowledge. This is done through various T&L strategies for teachers to adopt in order to achieve the objective of the lesson.

Assessment

Refers to the ways in determining the performance of the learner in learning by using various assessment tools

Time Allocation

Specified period of time (allocated hours per year) for teaching and learning to take place within school time table according to the scope of content determined in the syllabus at every level of schooling.

Necessary points to consider

Veracity of data

There is a need to point out that big data has its own implication especially on the aspect of the veracity of data in the modern age of IR4.0. The question of accuracy, trustworthy and precision are always a major concern when dealing with large data sets. Large amount of time is used to clean the data whether it involves structured or unstructured data. Therefore it is imperative to address this issue to the learners that they have to be very mindful on the implications and effects in making decisions based on analyzed data. Proper procedure has to be formulated in carrying out big data processing and analysis in order to achieve prudency in the findings and eliminate negative ripple effects in decision making that could create destructions.

Humanity & Values

Hence, the value of humanity, ethics, integrity, truthfulness, honesty and responsibility are the essence of important traits to educate learners towards future well-being and balance to the disruptive nature created by AI and Big Data in Industrial Revolution 4.0.

Suitability of Students Cognitive Level

It is important to develop sound curriculum content and learning resources that are suitable for the learners' cognitive levels and abilities with different learning behavior. The scope of learning at each level of schooling has to be structured in order to achieve meaningful learning experience to the learner.

Teacher education

According to Lew (2020), South Korea has developed a Software Teacher Education program focused on:

Securing the software education leadership capacity of elementary school teachers in preparation for the mandatory software education of the 2015 revised curriculum

Reinforcement of the teaching ability of elementary school teachers to foster creative and interdisciplinary talents of students pursued by the 2015 revision of the curriculum

Improving the creative and logical thinking power of elementary school teachers by improving their computing power.

Korea has proposed a curriculum with a very detailed list of teachers' circles that can be seen in (Lew, 2020).

Contents for High School

Algorithmic Thinking

> Instructions, conditionals, arrays, lists, cycles, decomposition, factorization, recursion, pseudo code, debugging.

Computational Modeling

> Arrays, dynamic systems, iteration, boards, agents, visualization, convergence, equilibrium, stability.

Machine Learning

> Training and test data sets, features, classes, scatter graphs, linear discriminators, decision trees, neural networks, genetic algorithm.

5. EXEMPLARS

Exemplars are examples of best practice which are designed to assist learners to increase their understanding of particular skills, content or knowledge in a given situation. They are tools that help to articulate established criteria and standards in the new curriculum. By using exemplars, it is a way to clearly explain and realize the curriculum into teaching and learning practice that can be understood and carried out by teachers. Exemplars are often used particularly during training or orientation courses in disseminating the curriculum.

Exemplar 1: Algorithmic thinking with coloring book for first grade

> Coloring books are very attractive to students and they can be a very powerful tool for learning. We show here some activities from (Araya, 2019) for understanding logical instructions with list, arrays, and searching procedures, and also for writing such type of instructions. These instructions are a kind of pseudo code, and have been designed as a preparation for developing algorithmic thinking through reading and writing pseudo code.
>
> Pages A of the coloring book (Figure 21) give instructions on what objects to paint and with

what colors. The corresponding page B has the same drawings of page A but the instructions are left blank and the student has to write new instructions for a classmate. The classmate will receive the sheet and has to paint according to the instructions. Thus page A is a model of pseudo codes ant on page B the student creates her own pseudo code.

Figure 21*: Colorea Ideas* coloring book and first graders writing coloring instructions for classmates

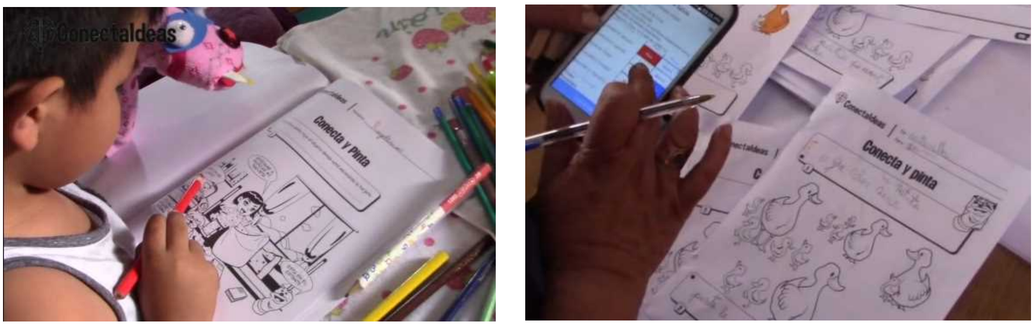


Figure 22: A first grader reading instructions and the teacher grading in her smartphone.

As shown in figure 22 at the right, the activities of all students are graded by the teacher in her smartphone (Araya, 2020a, Araya 2020b), each one associated with the national curriculum.

In figure 23 at the left is shown a page *A* where the first instruction is very direct, but the second one introduces a series of boxes each one with balls. This instructions introduces the notion of lists and arrays. At right is page *B* where the student has to write instructions (code) for a classmate.

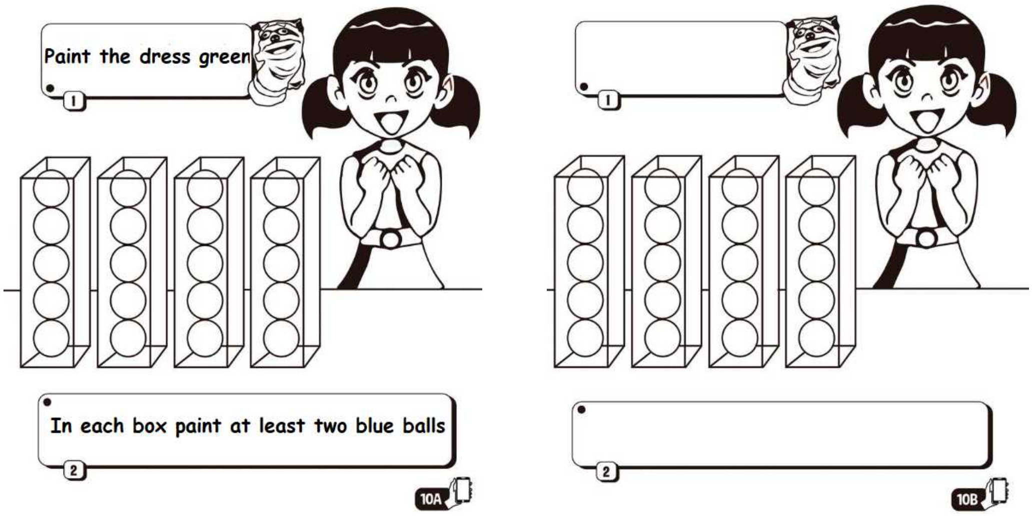


Figure 23: Pages A and B of the coloring book

Examples of possible instructions are:

- *For each box paint one ball more than in the neighboring box on its right*
- *While there is a box with all its ball white then paint red one white ball of that box*

Or more close to a typical computer code, the instruction could be:

- *For box 1 up to box 4 do*
    - *For ball 1 up to ball 5*
        - *Paint the ball with color red*
    - *End*
- *End*

Figure 24 at the left shows page *A*, a search exercise. This is a typical action in data bases. It requires reading a list of instructions given by different characters and infer where the object is hidden (the hidden object is never shown). At right is page *B,* where the student has to hide an imaginary object and write the instructions (code) to a classmate for searching and finding the object.

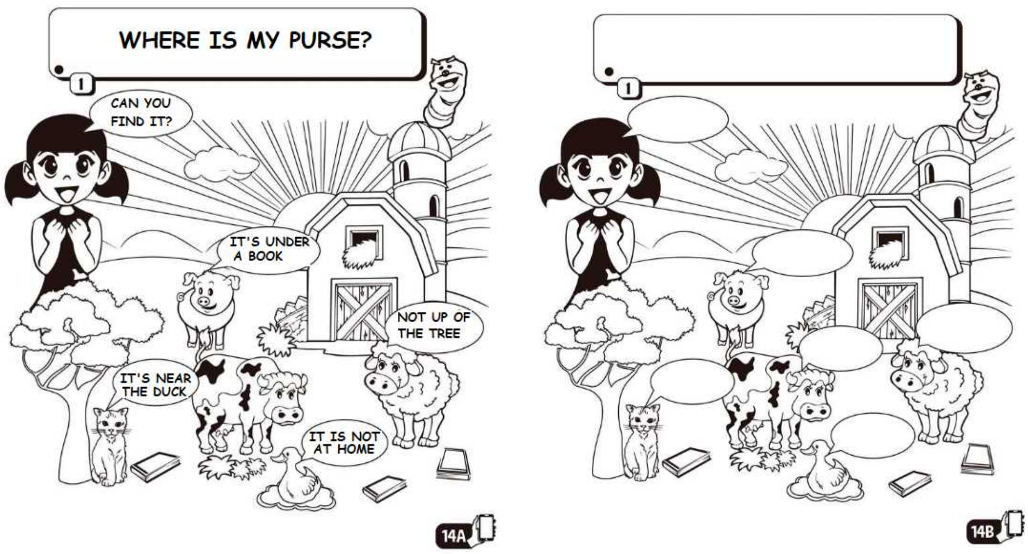


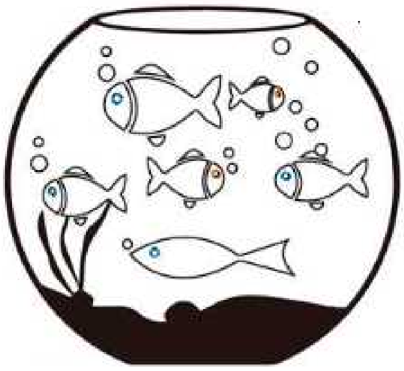
Figure 24: Search instructions on page A at the left. On page B students should write instructions to be decoded by a classmate.

Writing meaningful instructions that provides a unique place where the object is hidden is not a simple task. It requires careful planning and understanding, and later very probably a careful debugging of the instructions.

Exemplar 2: Machine learning with coloring book for elementary school students

How would you train a computer? Is it similar to train your dog? First, you need to show some examples to the machine. Then, you have to verify if the machine has really learned. This means that you need a set of examples for training, and another set of examples for testing. It is important that the testing examples are different examples from the ones used for training. Moreover, it is critical that the testing and training examples are selected randomly from the set of all examples.

What is an example? Imagine you are the trainer and one of your classmates is the machine. You will train him with the two fish bowls shown in figure 25. This is a sample with 10 fishes. The fishes have different features: sizes, directions, and eye colors. The task is to learn if size or direction is related to eyes colors. Then you have 10 examples. Your classmate have to figure out possible rules or patterns, and in a new fish bowl (a test set) he will have to paint the eyes of the fishes. Then you will compare with the colors of the test set and compute the agreement between his painting and the ones in the test fish bowl.



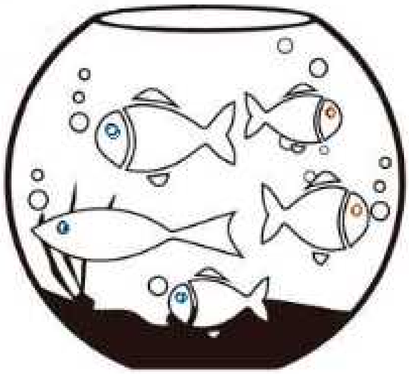


Figure 25: Two training fish bowls.

Figure 26 has the test with 4 fishes, but in this case the color of the eyes is not shown. Your classmate has to paint them.

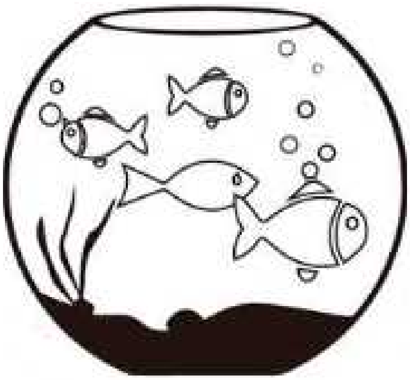


Figure 26: The testing fish bowl.

Once your classmate has painted them you compare with the colors that comes in the test set and then count in how many cases his production was correct. This will be the performance metric. You can compare with the predictions proposed by other students. Figure 27 has the test set with the true color of the fishes' eyes.

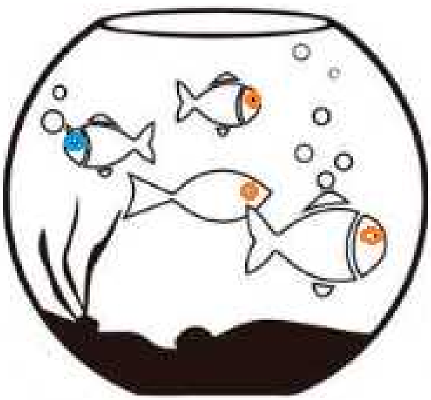


Figure 27: The fishes of the testing fish bowl with the true color of their eyes.

It is important that the students understand the notion of an example, a training set, a test set, the random selection of training and test set, and a performance metric. It is also important to have extra test set and measure the robustness of the predictions. It is also important that the students realize that if the training set is biased, then the machine will learn and produce biases solutions. ML is statistical thinking, and its connections to probabilistic intuitions are very important. As proposed in (Visotskiy, 2020) students can create task and simulate in electronic spreadsheets, in order to practice and improve their statistical and probabilistic intuitions.

Exemplar 3: Truck engine diagnosis

Figure 28: Mining track and an oil sample from the engine.

In mining, large trucks are critical. Therefore, a sample of engine oil is taken from every truck every 3 days and an analysis is performed (figure 28). One of the problems is to detect as early as possible water contamination. If so, the truck should be immediately sent for maintenance.

Table lists samples from 16 cases from trucks from the Chuquicamata copper mine (Reynolds et al, 1995). In 9 cases, after opening the engine, maintenance personnel found that the engine was in a normal condition. In 7 cases they found that the engine had water contamination. In the lab analysis of the engine oil, the lab technicians measured the particles per million of Fe, Cu, Na and they also measured the flash point of the oil.

The challenge is to develop a program or bot that reads the oil data and without stopping the truck and opening the engine, and automatically predicts whether the engine is normal or with water contamination.

Case	Fe	Cu	Na	Flash point	Diagnosis
1	19	3	6	204	normal
2	45	11	8	204	normal
3	28	15	8	222	normal
4	10	4	6	220	normal
5	39	15	17	204	normal
6	37	13	15	204	normal
7	17	4	8	226	normal
8	10	3	8	234	normal
9	24	39	15	212	normal
10	64	10	254	220	Water contamination
11	57	13	87	196	Water contamination
12	34	87	101	230	Water contamination
13	54	8	65	220	Water contamination
14	43	15	54	204	Water contamination
15	69	15	74	228	Water contamination
16	77	17	32	230	Water contamination

Table 2: 16 samples of oil, each one with 4 variables, and its diagnosis.

There are 6 possible scatter plots as shown in figure 29.

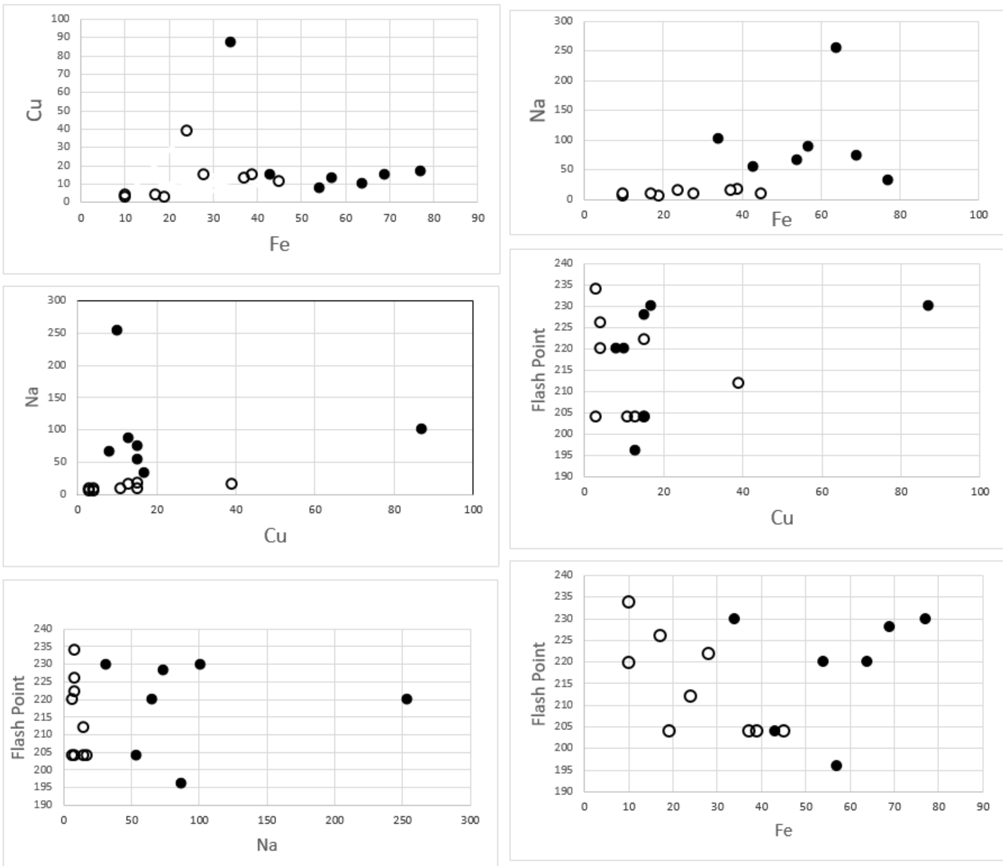


Figure 29: Six scatter plots. Each one with a pair of variables.

If we select the third scatter plot (figure 30) then with a ruler any student can draw a line separating black versus white dots, as shown in the next figure:

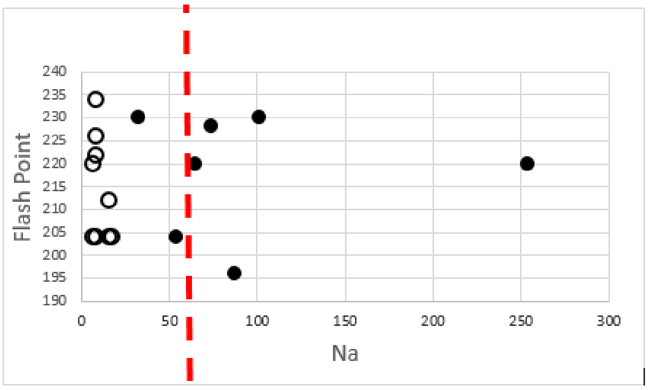


Figure 30: Scatter plot with Na concentration and Flash point.

This, this red line drawing can be translated in a decision tree, which has been learned by this simple Machine Learning Algorithm.

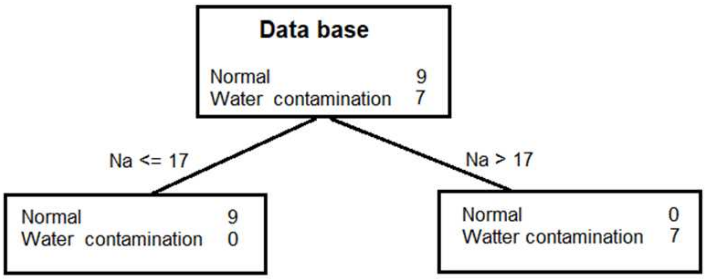


Figure 31:

This means that the data base has 16 cases (as shown in the top of figure 31), where 9 of them were normal and 7 of them had Water contamination. However on those cases with Sodium (Na) concentration less than or equal to 17 particles per million (ppm) all the 9 cases were normal. And, those cases with Na concentration higher than 17 ppm all of them had water contamination.

This exemplar promotes students to develop an attitude of searching for meaningful patterns, conceiving and exploring many symptoms and distinguishing them from diagnoses, and looking for empirical evidence of interconnections in the historical data.

Exemplar 4: Understanding of the outbreak of 2019 novel coronavirus (2019-nCoV).

World Health Organization

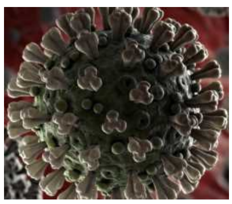


Figure 32: WHO COVID-19

Coronavirus is the world's most urgent problem today as shown in the WHO image in figure 32. This exemplar calls students to analyze it and contribute to finding solutions. It is of great significance for everyone that ML can help us in an emergency like this.

As shown in figure 33, up to February 9th , there have been:

- 814 deaths
- 37,595 cases
- 6,196 cases in severe conditions

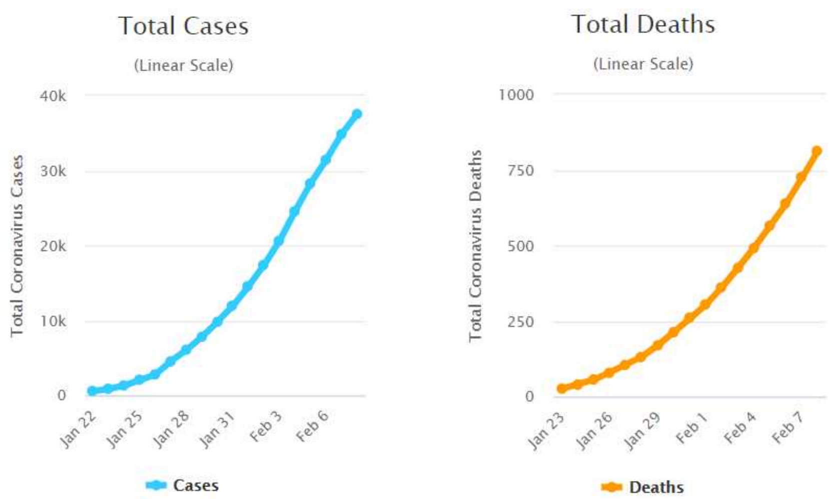


Figure 33: Total cases and total deaths

Figure 34 has the daily cases.

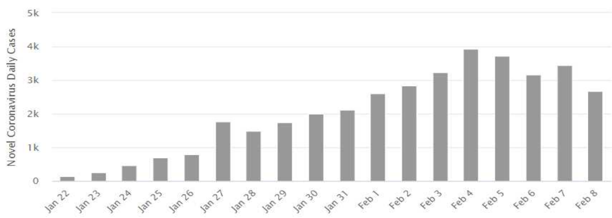


Figure 34: Daily cases

And Figure35 shows the daily cases growth factor.

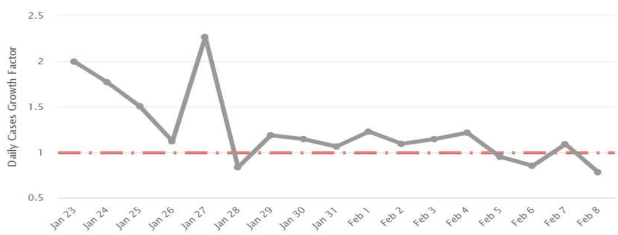


Figure 35: Growth Factor

What is the relation between these graphs? Figure 36 has the daily deaths

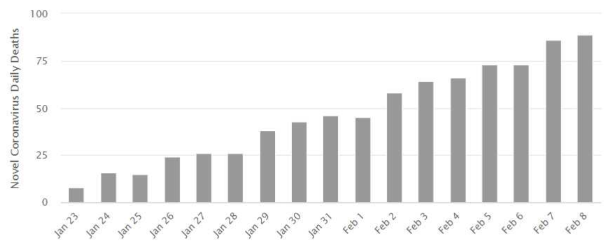


Figure 36: Daily deaths

And the daily deaths growth factor is in figure 37

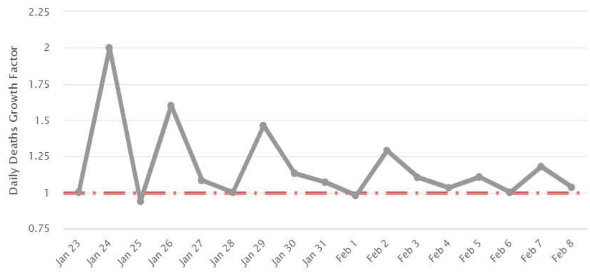


Figure 37: Growth factor of daily deaths

What are the relations between these graphs, and wat can of prediction can they help to do?

Consider now the updated data up to February 25th with 80,249 cases and 2,707 deaths.

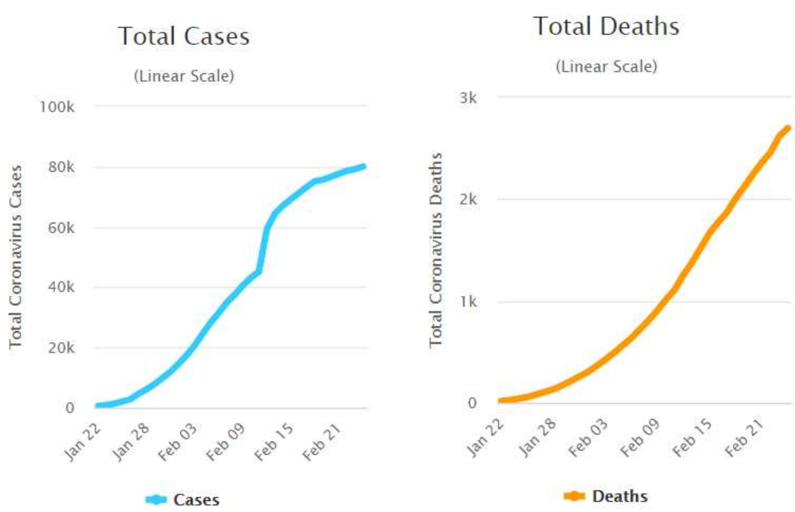


Figure 38: updated data up to February 21

Some interesting questions are:

- Were the previous predictions correct?
- Why the total cases graph has a break in the trends by mid-February?
- Can you correct the daily cases trends based on the daily deaths graphs?
- Can you make new predictions with this new updated data?

One critical question is to understand what are the characteristics of patients that can cause death due to infection of the Coronavirus? The first 17 deaths are in the following table (simplified version of Wang, Tang Wei, 2020, Updated understanding of the outbreak of 2019 novel coronavirus (2019-nCoV) in Wuhan, China. Journal of Virology. 29 January 2020 https://doi.org/10.1002/jmv.25689

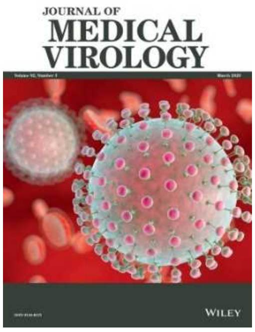


Figure 39: The Medical Virology Journal

If the information of these cases is obtained in detail as in the previous table with the first 17 cases, then two **Machine Learning** strategies can be done.

- **Death** versus **No Death**
- **Severe Condition** versus **Non Severe Condition**

With the first data base Machine learning can be applied to detect features or combination of features that will predict deaths. With the second data base Machine Learning can help find features or combination of features that will predict severe condition.

Case	Gender	Age	1st sympt	Comorbidity	Surgery	1st sympt to death, days
1	M	61	Fever	yes	NA	20
2	M	69	Fever	yes	NA	16
3	M	89	No Fever	yes	NA	10
4	M	89	No Fever	yes	Yes	6
5	M	66	Fever	yes	Yes	10
6	M	75	Fever	yes	Yes	14
7	F	48	Fever	yes	NA	41
8	M	82	No Fever	NA	NA	12
9	M	66	No Fever	NA	NA	30
10	M	81	Fever	NA	NA	11
11	F	82	Fever	yes	NA	19
12	M	65	No Fever	NA	NA	13
13	F	80	Fever	yes	NA	11
14	M	53	Fever	NA	NA	20
15	M	86	No Fever	yes	Yes	19
16	F	70	Fever	NA	NA	8
17	M	84	Fever	yes	Yes	16

Table 3: 1st symptom to death

Another critical question is to find the characteristic of the patients that produce a much faster death. From the data already available for the first 17 patients that died from Coronavirus infection (table 3), we can compute the mean of days of first symptom to death, and then classify as fast deaths those patients that died before that average.

Table 4 has an extra column at the extreme right that marks with one those patients with fast death (first symptom to death less than average), and zero those with sower deaths (first symptom to death higher or equal than average).

Case	Gender	Age	1st sympt	Comorbidity	Surgery	1st sympt to death, days	1st sympt to death, above average?
3	M	89	No Fever	yes	NA	10	0
4	M	89	No Fever	yes	Yes	6	0
6	M	75	Fever	yes	Yes	14	0
8	M	82	No Fever	NA	NA	12	0
10	M	81	Fever	NA	NA	11	0
11	F	82	Fever	yes	NA	19	1
13	F	80	Fever	yes	NA	11	0
15	M	86	No Fever	yes	Yes	19	1
16	F	70	Fever	NA	NA	8	0
17	M	84	Fever	yes	Yes	16	1
1	M	61	Fever	yes	NA	20	1
2	M	69	Fever	yes	NA	16	1
5	M	66	Fever	yes	Yes	10	0
7	F	48	Fever	yes	NA	41	1
9	M	66	No Fever	NA	NA	30	1
12	M	65	No Fever	NA	NA	13	0
14	M	53	Fever	NA	NA	20	1

Table 4: 1st symptom to death above average

If we plot on a number line the age of each patient but marking with a circle those with slow death and with a cross those with fast death, then we have the following one dimensional scatter graph (figure 40);

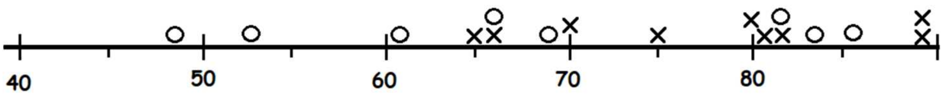


Figure 40: On dimension scatter plot

From this one dimensional scatter graph we can see a clear pattern. Older patients die faster. But we can also identify the critical age, and represent it with a decision tree (figure 41):

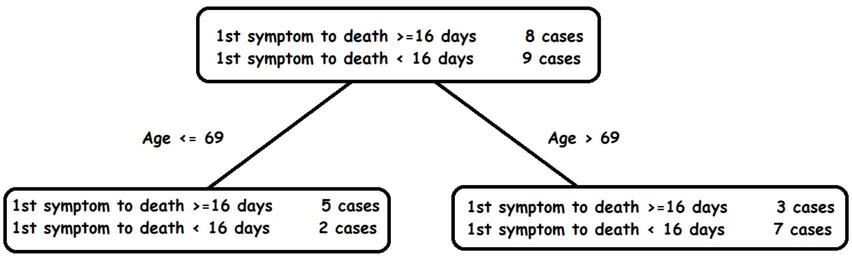


Figure 41: Decision tree induced from the data

ACKNOWLEDGEMENT

Support from ANID/ PIA/ Basal Funds for Centers of Excellence FB0003 is gratefully acknowledged, and from María Jesús Honorato from the Curriculum Unit of the Ministry of Education of Chile.

REFERENCES

1. Araya, R. (2020a) Mobile Performance Support System for Teachers and Parents Teaching First Graders to Read. 16th International Conference on Mobile Learning proceedings.
2. Araya, R. (2020b) A territorial learning ecosystem for parents' participation and cooperation. 5th conference on Smart Learning Ecosystems and Regional Development SLERD 2020.
3. Araya, R., Isoda, M., González, O. Inprasitha, M. (2019) APEC Project InMside Discussion Document for Computational Thinking. A Framework for Computational Thinking for the Transition to a Super Smart Society
4. Araya, R. & Gigon, P. (1992) Segmentation Trees: a New Help for Building Expert Systems and Neural Networks" Computational Statistics, COMPSTAT 92, 1:119-124. Physica-Verlag HD. http://link.springer.com/chapter/10.1007/978-3-662-26811-7_17#
5. Araya, R. (2017) Clases Públicas STEM Incendios Forestales. CIAE. https://www.conectastem.cl/conecta/Libro/samples/Libros/
6. Araya, R. (submitted) Math Teacher Education to Prepare Students for the 21st Century Jobs: Mathematical Modeling and Computational Thinking
7. Araya, R. (2019) *Colorea Ideas* Coloring Book. AutoMind. https://www.conectastem.cl/colorea-ideas/
8. Araya, R., Isoda, M., & González, O. (2020). A Framework for Computational Thinking in Preparation for Transitioning to a Super Smart Society. Journal of Southeast Asian Education, (1), 1-15. Chi Thanh (2020) Computational Thinking: Concept and example in Vietnamese new curriculum. Downloaded from http://www.criced.tsukuba.ac.jp/math/apec/apec2020/presentations/12Feb/15-Nguyen_C.T-2020.pdf
9. DeLanda, M. (2017) The Use of Genetic Algorithms in Art.
10. González, O., Isoda, M., Araya, R., Inprasitha, M. (2019) APEC Project InMside Discussion Document for Statistical Thinking.
11. Isoda, M. (2020) The 37th HRDWG-EDNET Meeting. Downloaded from http://www.criced.tsukuba.ac.jp/math/apec/apec2020/presentations/11Feb/01-ISODA_M-2020-1.pdf

12. Isoda, M.; Katagiri, S. Mathematical Thinking. How to Develop it in the Classroom. Word Scientific.(2012)
13. Jordan, M, (2013) On Statistics, Computation and Scalability. Bernoulli 19(4), 2013, 1378–1390
14. Jordan, M. (2019) Artificial Intelligence—The Revolution Hasn't Happened Yet. https://hdsr.mitpress.mit.edu/pub/wot7mkc1/
15. Kano, T. (2020) Japanese Software Education. Downloaded from http://www.criced.tsukuba.ac.jp/math/apec/apec2020/presentations/12Feb/12-Toshiharu_K-2020.pdf
16. Kenyon, G. (2019)　AI's Big Challenge. Scientific American.
17. Koshi, U. (2020) Information education and teacher training at junior high school. Downloaded from http://www.criced.tsukuba.ac.jp/math/apec/apec2020/presentations/11Feb/03-Koshi_U-2020.pdf
18. Lambas, S. (2020) Problem Solving Model for analyzing data in digital era for Senior High School: "big picture with incomplete detail for some components"
19. Lew, H. (2020) AI for Education in Korea: Case of KNUE. Downloaded from http://www.criced.tsukuba.ac.jp/math/apec/apec2020/presentations/11Feb/07-Hee-chan_L-2020.pdf
20. Ministry of Education Malaysia (2020) Computational Thinking In Primary and Secondary School Curriculum. Downloaded from http://www.criced.tsukuba.ac.jp/math/apec/apec2020/presentations/12Feb/13-Mash_M-2020.pdf
21. Reynolds, A. & Araya, R. (1995) Building Multimedia Performance Support Systems. McGraw-Hill, New York.
22. Santillan, M (2020) InMside II project. Downloaded from http://www.criced.tsukuba.ac.jp/math/apec/apec2020/presentations/11Feb/09-Marcela_S-2020.pdf
23. Siegler, R. S., & Araya, R. (2005). A computational model of conscious and unconscious strategy discovery. In R. V. Kail (Ed.), Advances in child development and behavior, Vol. 33 (pp. 1-42). Oxford, UK: Elsevier
24. TajulArus, S. (2020) Knowledge and Skillsets for Curriculum IR4.0. Downloaded from http://www.criced.tsukuba.ac.jp/math/apec/apec2020/presentations/11Feb/04-Sofian_T-

2020.pdf

25. Thipakorn, B. (2020) Computational Thinking and Humanity for 4IR Downloaded from http://www.criced.tsukuba.ac.jp/math/apec/apec2020/presentations/11Feb/02-Bundit_T-2020.pdf

26. Tin Lam, T. (2020) Computational Thinking. Downloaded from: http://www.criced.tsukuba.ac.jp/math/apec/apec2020/presentations/11Feb/06-Toh_TL-2020.pdf

27. Ulep, S. (2020) Computational Thinking in Junior High School, Downloaded from http://www.criced.tsukuba.ac.jp/math/apec/apec2020/presentations/12Feb/17-Soledad_A.U-2020.pdf

28. UNESCO (2019) BEIJING CONSENSUS on artificial intelligence and education

29. Visotskiy, I. (2020) The Collector's Problem: An Example of Inciting to Thinking. Downloaded from http://www.criced.tsukuba.ac.jp/math/apec/apec2020/presentations/11Feb/08-Vysotsky_I-2020.pdf

30. Wang, Tang Wei, 2020, Updated understanding of the outbreak of 2019 novel coronavirus (2019-nCoV) in Wuhan, China. Journal of Virology. 29 January 2020 https://doi.org/10.1002/jmv.25689

31. Wing, J. M. (2006) Computational thinking. Commun. ACM, 49, 33–35.

32. Wylie, C. (2018) How 300 Matchboxes Learned to Play Tic-Tac-Toe Using MENACE https://opendatascience.com/menace-donald-michie-tic-tac-toe-machine-learning/

33. Yunianto, W. (2020) Teacher Training On: Programming & Computational Thinking http://www.criced.tsukuba.ac.jp/math/apec/apec2020/presentations/12Feb/16-Wahid_Y-2020.pdf

Chapter 3.

Recommendations for Curriculum Reform on Statistical Thinking

Orlando González	Masami Isoda	Roberto Araya
Assumption University, Thailand	University of Tsukuba, Japan	University of Chile, Chile

In this chapter we propose recommendations and exemplars for a curriculum on Computational Thinking for the Digital Economy (Isoda, 2020). Statistical Thinking has been enhanced since 1990's and here we discuss how different is the statistical thinking for big data and AI. We will discuss our views of statistical thinking in relation to Big Data and statistical thinking in relation to the subjective probability running behind the AI system and them, conclude recommendations.

INTRODUCTION: STATISTICAL THINKING ON THE ERA of BIG DATA

The interest in big data is growing exponentially in today's society. Commercial insights, government initiatives and even research's calls, all seem to be focused on exploiting the potential of technology to capture and analyze massive amounts of data in increasingly powerful ways. Big data, that is, data that are too big for standard database software to process, is everywhere. For some, big data represents a paradigm shift in the ways that we understand and study our world, and at the very least it is seen as a way to better utilize and creatively analyze information for public and private benefit.

The concept of big data "refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze" (Manyika et al., 2011, p.1). Additionally, big data is often associated with key characteristics that go beyond the question of size, namely the 5 Vs: volume, velocity, variety, veracity and value (Storey & Song, 2017). Big data is dispersed among various platforms that operate with different standards, providers and degrees of access (Ferguson, 2012). For example, a lot of work in big data focuses on Twitter, the blogosphere, and search engine queries. All of these activities are not undertaken equally by the whole population, which raises concerning issues around the question of whose data traces will be analyzed using big data.

There are also a number of practical issues related to working with big data. These include, among others, issues we cannot afford to ignore, such as implications for the training of future teachers regarding handling and analysis of big data.

Due to the fact that big data has recently become mainstream in many research fields, including education, it is important to discuss and answer the following questions:

In order to function effectively in a society driven by big data and digital economy, what are the necessary processes of statistical thinking required to handle big data?

How can we revise current curriculum frameworks of statistical thinking to incorporate big data for the digital economy?

How can we incorporate core ideas of big data for the digital economy into the high school curriculum? What are plausible instructional activities (exemplar applications) for teaching the fundamental ideas of statistics while fostering statistical thinking for big data and the digital economy?

1. NATURE OF BIG DATA

Big Data is produced by Data Science which is not the same as Statistics on the Neyman-Pearson context. It is not necessary that data is given to begin statistical analysis and inference. On the Data Science, it produces the new data by connecting no related data statistically and analyzed it by using mathematics which includes Bayesian Statistics with intuitive probability. The followings are the 5Vs to explain the nature of big data.

Volume: This refers to data at rest. Datasets with sizes in the order of terabytes, petabytes and zettabytes.

Variety: This refers to the many forms (e.g., text, images, videos, audio files, emails) and sources (e.g., spreadsheets, databases, social media and monitoring devices) of data.

Velocity: This refers to data in motion: the speed at which data flows in from sources (e.g., real time streaming).

Veracity: This refers to data in doubt. Depending on its origin, processing technologies, and collection methods, data can have biases and inaccuracies attached, which need to be identified and accounted for.

Value: This refers to turning data into profit. Just having big data is of no use unless we can turn it into value**.**

On the result of data science, currently maps are used for driving and shopping. Even though numerical data is running and connected behind, the interface does not show it. However users are indirectly using it by operating visible and manipulative interface. Even though different data is connected, if users could not recognize their reasoning is based on biased data, they believe the result without any doubt. Users of Big Data must have, in this context, statistics education which is necessary for all users of BIG Data to develop the critical skills such as worry questions (Gal, I., .2002). However, statistical thinking is not limited to critical skills to doubt the using of Big data on the era of AI. What are the necessary components of statistical thinking on the era?

2. FRAMEWORK FOR STATISTICAL THINKING

Formal Neyman-Pearson Statistics usually applies formal statistical models/theories for given data sets such as statistical tests by using Normal Distribution. Exploratory Data Analysis proposed by Tukey, J.W., a novel economy medalist, prefers the statistics for analyzing data sets to summarize their main characteristics with visible representations to know what the data can tell us beyond the formal modeling or hypothesis testing task. These statistical approaches are the methods of analysis for the data sets. However, data sets on the data science are unstable to apply Neyman-Pearson Statistics because they are not fixed data sets. Data sets of Big Data usually change such as continuously adding, seeking, mining and connecting. This nature extends statistical thinking from the current statistics at schools to unknown era. For example, Bayesian Statistics is not currently treated at the high school level but are usually used as an alternative approach for Neyman-Pearson Statistics on the context of Data Sciences. By considering the current issues, we would like to propose the following curriculum reform framework for Statistical Thinking (Figure 1).

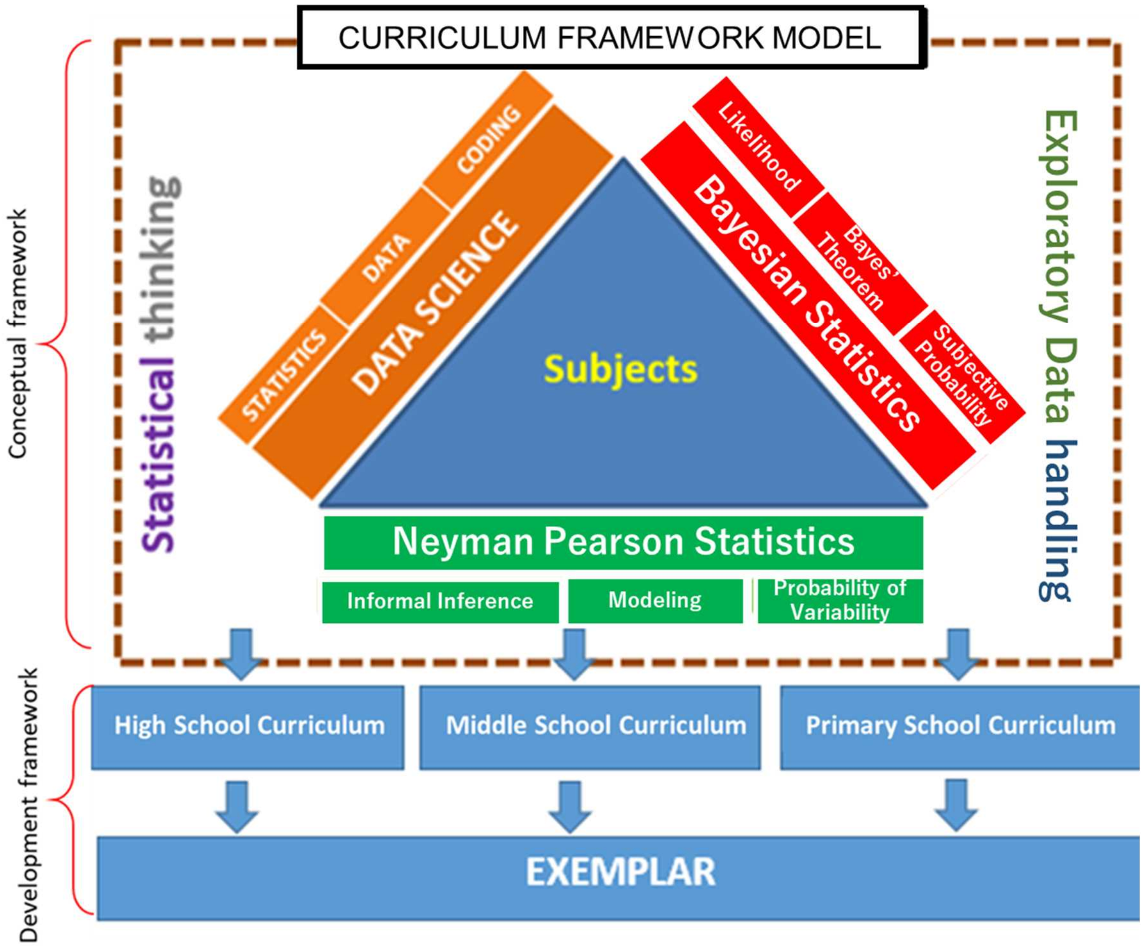


Figure 1. Curriculum Framework for Statistical Thinking on InMside

On figure 1, any content of subject can be seen from the perspectives of Data Sciences, Neyman-Pearson Statistics and Bayesian Statistics. Thus, any school subject can be reconsidered from these perspective. Even though figure 1 can be applied any school subject, in the following discussion, we mainly focus on the statistics, mathematics and informatics curriculum.

In relation to current existing school statistics curriculum, many researchers (e.g., Wild & Pfannkuch, f1999; adelMas, 2002; Watson, 2017) consider statistical thinking as the practice of statistics through the enactment of the different thought processes involved in statistical problem solving and statistical investigations. For us, in this digital era, statistical thinking processes do not follow the Problem-Plan-Data-Analysis-Conclusion (PPDAC) cycle (Wild & Pfannkuch, 1999) anymore due to the shift in the way we work with data sets with the arrival of big data analytics. In fact, the PPDAC cycle is a question-then-answer research method, focused on gathering data through planned processes with a purpose, chosen on statistical grounds to justify certain types of inferences and conclusions. However, in times of big data, this is actually a weakness of the PPDAC cycle because most of the available data are opportunistic (happenstance or "found") data (including "big data"): huge amounts of data already collected by others and hosted somewhere. Nowadays, many companies have data teams exploring large sets of raw opportunistic data, looking for new connections and identifying significant correlations, while refining their analysis until they arrive at valuable understandings. This approach reverses the question-then-answer process of the PPDAC cycle. It starts with strong data-first for answering question and then works backwards to find the questions that should have been asked. PPDAC cycle also includes Exploratory Data Analysis in the process of Analysis.

On statistics education research, there are up-to-date statistical thinking process designed to consider these criticisms to the PPDAC cycle. By doing so, in the context of using Big Data which has the nature of 5Vs, we came up with the following key phases on handling big data for the statistical thinking process to describe how a person engages in statistical thinking. The proposed key phases explain statistical thinking process for handling Big Data as a cognitive process which are comprised of the following five phases:

Patterns and relationships from data: Look for patterns and relationships within the data based on a particular interest.

Questions: Pose critical and worry questions, in order to find plausible explanations to the patterns and relationships found.

Objectives: Set objectives related to the posed questions, in order to analyze the data.

Data mining: Re-examine the data according to the objectives, explore the old and new data sources, or introduce new variables for consideration. Data mining can be data-oriented, explanation-oriented, or future-oriented.

Understanding and/or designing: Provide ideas for new activities based on the understanding of the past, and design plans and strategies for the future based on the results from the data mining.

These phases are not necessary in this order like Exploratory Thinking process in the PPDAC Cycle.

In order to describe figure 1, we illustrate two exemplars to explain the phases for exploring Big Data and the ways of Bayesian Statistical Approach in the following sections. We do not illustrate Neyman-Pearson Statistical Approach here because it currently exists in the school curriculum.

1. EXEMPLAR APPLICATION OF THE FRAMEWORK: AGING POPULATION ISSUES

IN APEC MEMBER ECONOMIES

For the purpose of exemplifying this framework, let us suppose that we are interested in exploring issues related to aging population, which is a concerning issue in many societies, such as the Japanese. So, we may check the worldwide trend of web searches for terms such as "social security" and "nursing home", focusing on APEC economies.

For this example, we will use the website "Google Trends" (https://trends.google.com), which is an open-access online platform for big data that enables creative discovery from information on how frequent a given search term was entered into Google's search engine in real-time or within given time and date constraints.

Patterns and relationships from data

Using "Google Trends", we looked for patterns and relationships within the data hosted in the platform, based on our interest in aging population issues on APEC member economies. Then, by checking the worldwide trend of web searches for terms related to aging population, such as "social security" and "nursing home", we will be able to identify possible patterns and relationships in the data regarding these terms. Figure 2 shows the evolution of the worldwide search trends for "social security" and "nursing home" in the past 15 years.

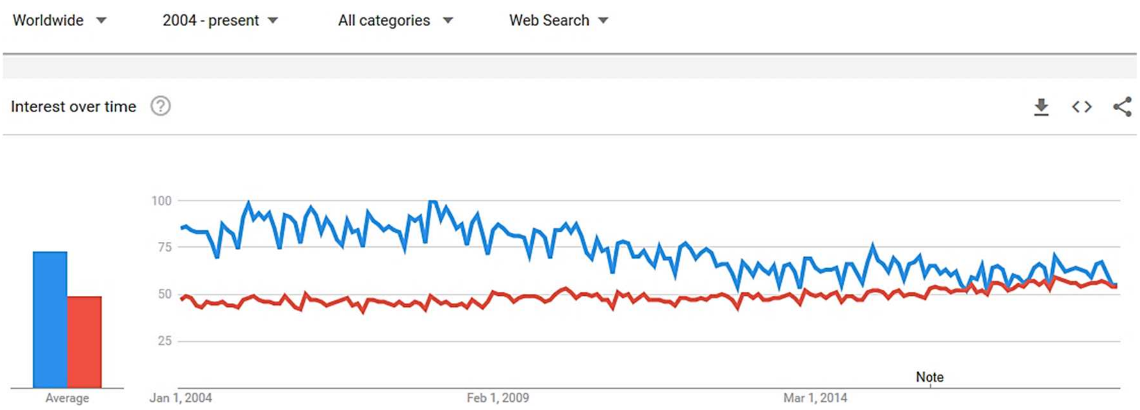


Figure 2: Evolution of the worldwide Google search trends for "social security" (in blue) and "nursing home" (in red) in the past 15 years.

Now, let us focus on the online search trends for the terms "social security" and "nursing home" focusing on APEC member economies. In the top row of Figure 3, we can see that Japan, Canada and the US were APEC member economies in which online searches for the term "nursing home" were higher in comparison to the term "social security" in the last year. On the other hand, in the bottom row of Figure 3, we can see that Chile, Peru and the South Korea were APEC member economies in which online searches for the term "social security" were higher in comparison to the term "nursing home" in the last year.

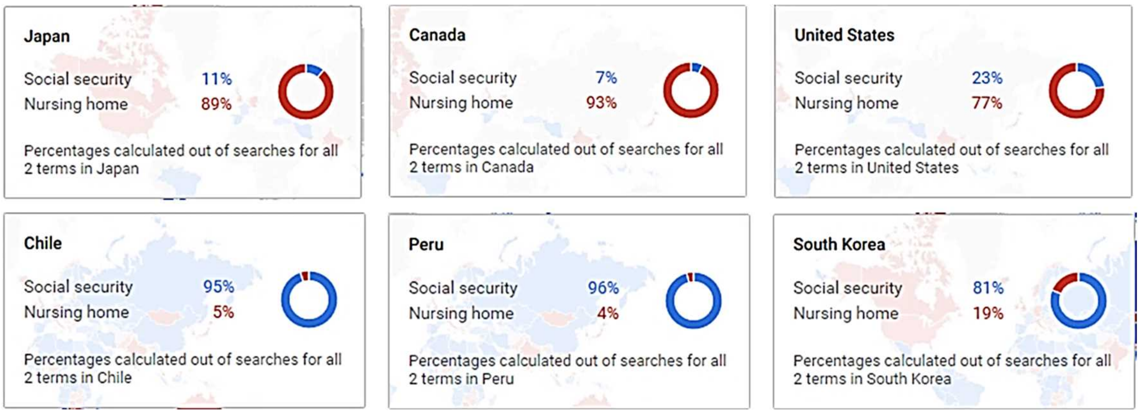


Figure 3: Percentage comparison of Google searches for the terms "social security" (in blue) and "nursing home" (in red) in six APEC member economies in 2018.

Questions

From the visual representations obtained by using Google Trends, it is possible to pose a series of questions regarding the statistical information being displayed, in order to find explanations to the patterns and relationships found. These questions can support the process of critical evaluation of statistical messages and lead to the creation of more data representations, informed interpretations and judgments.

Some questions that might be posed for this example are the following:

- Why do some member economies (such as Japan, Canada and USA) seem to show considerably more interest on "nursing home" than on "social security"?
- Why do some member economies (such as Chile, Peru and South Korea) seem to show considerably more interest on "social security" than on "nursing home"?
- What are the reasons for the decreasing and increasing trends of the graphs?
- What could be the behavior of these trends for individual APEC member economies (such as the ones mentioned in Figure 3) in the next decade?
- In member economies such as Japan, Canada and USA, where it seems to be more interested on "nursing home" than on "social security", what is the behavior of related queries, such as "nurse"?
- In member economies such as Chile, Peru and South Korea, where it seems to be more interested on "social security" than on "nursing home", what is the behavior of related queries, such as "tax" or "pension"?
- How is the social security policy in member economies showing considerably more interest on "nursing home" than on "social security" (such as Japan, Canada and USA)?
- How is the current state of nursing home services provided to the elderly in member economies showing considerably more interest on "social security" than on "nursing home" (such as Chile, Peru and South Korea)?

Objectives

Now, from the posed questions, we are able to set clear objectives. In fact, each objective should be associated to at least one question. In this example, some objectives stemming from the questions above are the following:

1. To look for and identify the reasons why some APEC member economies seem to show considerably more or less interest on “nursing home” than on “social security”.
2. To determine the behavior for individual APEC member economies regarding individual queries, such as the mentioned above.
3. To predict the trends of web searches for the terms “nursing home” and “social security” in APEC member economies in the next decade.
4. To determine the behavior of related queries (e.g., “nurse” for member economies showing more interest on “nursing home”, and “tax” or “pension” in member economies showing more interest on “social security”) in APEC member economies with a particular search trend.

Data mining

During this phase of the statistical thinking process, addressing the objectives set in the previous phase will lead to a re-examination of the data, from which new insights and knowledge discovery will emerge from three types of data mining: big data-oriented, explanation-oriented, and future-oriented data mining.

For the purpose of exemplifying this phase, let us address the Objective 1 (i.e., to look for and identify the reasons why some APEC member economies seem to show considerably more or less interest on “nursing home” than on “social security”). In the case of Japan and other APEC member economies, the main reason can be the current structure of the population pyramid (explanation-oriented data mining). In order to construct plausible explanations from the population pyramids, we need to select and transform the necessary data into the required form (big data-oriented data mining), as presented in Figure 4.

So, if we look at the population pyramids of the APEC member economies in the top row of Figure 3, we can see different structures among the member economies, regardless the fact that online searches in 2018 for the term “nursing home” were higher in Japan, Canada and the US, in comparison to the term “social security” (see Figure 4).

In Japan, we can observe a large proportion of the population are close to or over retirement age. In Canada, we can observe a large proportion of the population are close to retirement age. In the US, this phenomenon is not an issue, because a large proportion of the population are below 50 years-old, far from retirement age. Thus, plausible explanations of why some APEC member economies seem to show considerably more or less interest on “nursing home” than on “social security” may vary from economy to economy. In the cases of Japan and Canada, a large group of elderly people who are close to retirement, or already retired, might be planning to live in a nursing home. In the case of the US, young people, starting to make a live by themselves, might be looking information on nursing homes for their elderly parents.

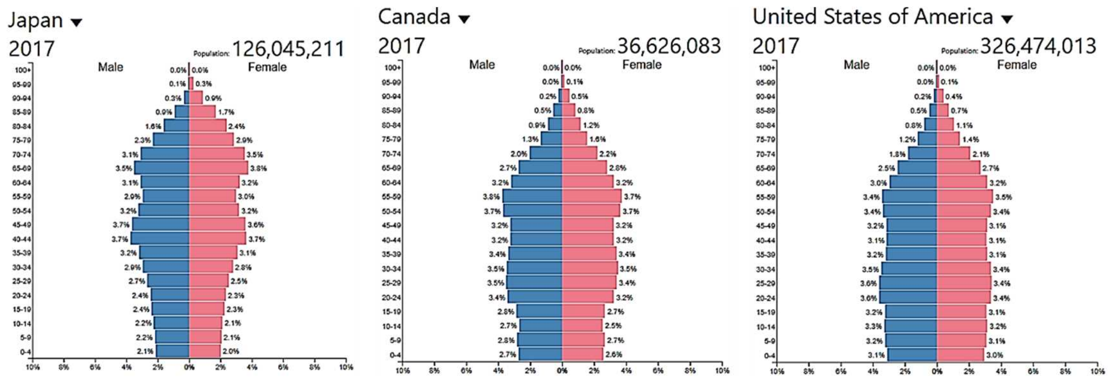


Figure 4: Population pyramids in 2017 for Japan, Canada and the USA.

From these hypotheses, we might make inferences, imagining the future of "nursing homes" in Japan, Canada and the US (future-oriented data mining), using big data to support our plausible explanations for the future to come (see Figure 5).

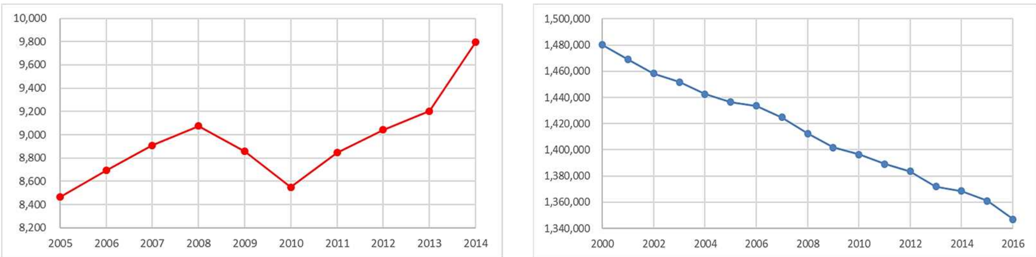


Figure 5: Number of nursing homes in Japan (red line, 2005–2014) and United States (blue line, 2000–2016).

Understanding and/or designing

From the discovered knowledge, inferences and plausible explanations generated in the previous phase, we have gained valuable understanding about the topics of interest ("nursing home" and "social security"). This understanding provides us with ideas to develop new activities related to the topics or variables of interest. Understanding the past allows us to design for the future, all supported on the data mining results.

In our exemplar application, from the understanding of rising need for nursing homes and nurses in Japan, someone could design business plans targeting senior citizens: in-home care services, senior citizen transportation services, e-commerce store for the elderly, wheelchair manufacturing, foreign nurse recruitment agency, and so on. In the USA, most of these ideas (e.g., a foreign nurse recruitment agency to provide care service for the elderly) will not work everywhere, but could be successful in states such as Oregon and Virginia, as shown in Figure 6.

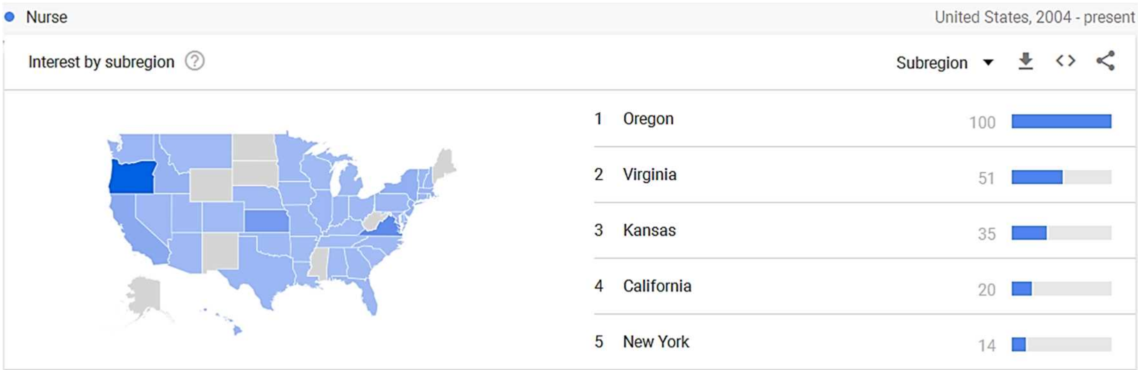


Figure 6: Interest in Google searches for the term "nurse" in United States in the last 15 years.

2. STATISTICAL THINKING FOR USING BIG DATA

In times of big data, artificial intelligence and digital economy, statistical thinking has evolved from a traditional question-then-answer analysis, through which we ask questions and then collect and analyze the data to arrive at a conclusion that can be used to make decisions. Now, we should use a more disruptive and creative approach, which starts with data-first answers from the examination of opportunistic data (which is data that just happen to be available in electronic form because they have accumulated for other reasons by other people), and then works backwards to find the questions that should have been asked.

Under this scenario, previous frameworks of statistical thinking, such as the PPDAC cycle, is not appropriate to explain the richness and complexity of thinking involved in real-world statistical investigations dealing with big data. On that regard, we developed a new framework for statistical thinking, comprised of five phases or cognitive processes (i.e., patterns and relationships from data, questions, objectives, data mining, and understanding and/or designing), in order to appropriately describe how a person engages in statistical thinking while handling big data. An exemplar application of the framework illustrated the richness and complexity of thinking involved in handling big data with Google Trends, exploring issues related to aging population, by checking the worldwide trend of web searches for terms such as "social security" and "nursing home", focusing on APEC member economies. From this application, it was possible to identify member economies with similar and contrasting characteristics, and classify APEC economies based on such characteristics.

Although we have illustrated our proposed new framework for statistical thinking with an exemplar application, issues regarding how we can incorporate core ideas of big data for the digital economy into the high school curriculum of each APEC economy, as well as what plausible instructional activities (exemplar applications) can be designed for teaching the fundamental ideas of statistics while fostering statistical thinking for big data and the digital economy, are expected to be fundamental outcomes from the discussion of the current document among all the participants to this APEC-Tsukuba International Conference 2019.

3. Exemplar of Bayesian Statistics for COVID-19

For explanation of a part Bayesian Statistics of the curriculum reform framework in Figure 1, we would like to illustrate the case of checking the disease of COVID-19 by using Bayesian Theorem (Suematsu, M., to appear). If we use Bayesian Theorem, we can consider the future by postulating the unknown probability as subjective even though the data is not pre-requested.

How Polymerase Chain Reaction (PCR) test is appropriate and meaningful to check COVID-19?

Events for positive test results are A, events for negative tests are B, events for disease (infected) are C, and events for not disease (uninfected) are D.

Sensitivity: The percentage of people who are sick and tested positive: $SEN = P(A|C)$

Specificity: Percentage of people who are not ill and tested negative: $SPE = P(B|D)$

When the test result is positive, the probability of being infected is PCR (T), and when the test result is negative, the probability of being uninfected is PCR (F).

If we apply this to Bayes' Theorem, we got the following:

$$PCR(T) = P(C|A)$$

$$= \frac{P(A|C)P(C)}{P(A)}$$

$$= \frac{P(A|C)P(C)}{P(A|C)P(C) + P(A|D)P(D)}$$

$$= \frac{SEN \times P(C)}{SEN \times P(C) + \left(1 - SPE\right) \times P(D)}$$

$$PCR(F) = P(D|B)$$

$$= \frac{P(B|D)P(D)}{P(B)}$$

$$= \frac{P(B|D)P(D)}{P(B|C)P(C) + P(B|D)P(D)}$$

$$= \frac{SPE \times P(D)}{(1 - SEN) \times P(C) + SPE \times P(D)}$$

Tentatively, we set the following condition.

$SEN = 0.7$, $SPE = 0.99$

And set the prior probability for events for infected is C: If we assume

$$P(C) = x,$$

Then, the prior probability for events for not infected is D.

$$P(D) = 1 - x$$

Then, we get the following graphs; Prior probability distribution (Figure 7)

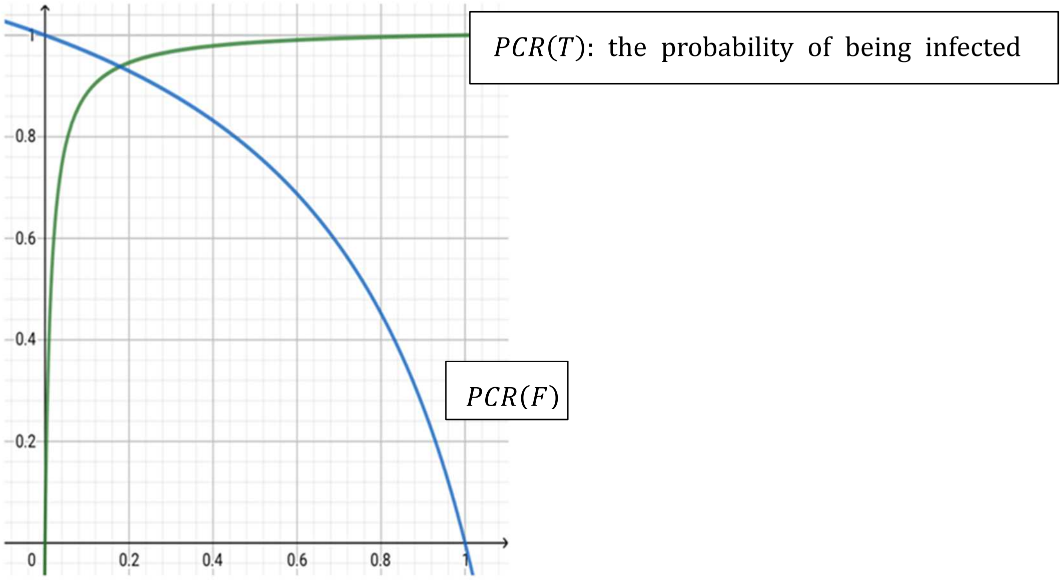


Figure 7. Graphs of Probabilities PCR (T) as Test Positive for diseased and PCR (F) as Test Negative for not diseased

Implication from the graphs (Figure 7)

PCR (T) deteriorates when the prior probability P (C) is low. For example, when P (C) = 0.1, the probability of being infected is less than 90% even if the test result is positive. In addition, PCR (F) deteriorates when the prior probability P (C) is high. For example, when P (C) = 0.9, the probability of being uninfected even if the test result is negative is about 25%.

Further analysis

False Test-Positive Rate P (D | A): Probability of being uninfected even with a positive test

False Test-Negative Rate P (C | B): Probability of being infected with a negative test

P (D | A): Positive test results must be isolated, filling up the number of beds even though they are not infected.

$$P(D|A) = 1 - P(C|A) = 1 - PCR(T)$$

P (C | B): If the result of the test is negative, the infection may spread to the surroundings without taking measures such as masks with confidence.

$$P(C|B) = 1 - P(D|B) = 1 - PCR(F)$$

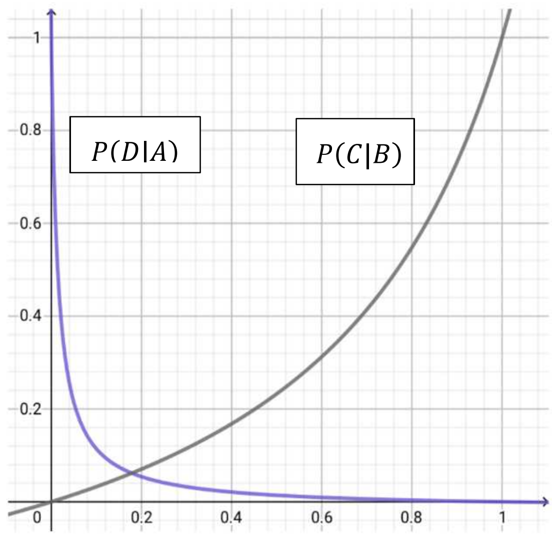


Figure 8. Graphs of the False Test-Positive Rate P (D | A) and False Test-Negative Rate P (C | B):

P (D | A) worsens when the prior probability P (C) is low. For example, when P (C) = 0.1, the probability of being infected is about 12% even if the test result is positive. In P (D | A), the prior probability P (C) is high and deteriorated. For example, when P (C) = 0.9, the probability of being infected is about 73% even if the test result is negative.

Neyman-Pearson Statistics used given data to analyze. Bayesian Statistics tentatively set the unknown probability and reason by using Bayesian Theorem. The probability is unknown yet reasoning is still possible. Neyman-Pearson statistical inference, statistical test, are done by distribution however Bayesian statistics are used for likelihood for comparison. The comparison (greater or less) of the values of likelihood is the way for making decision by decision making tree on AI (see Chapter 2, Exemplar 3 and 4). On the decision making on AI system, likelihood (or odds) can be obtained from probability functions such as the graphs like Figure 7 and 8, instead of statistical text based on statistical model distributions. It is very much different ways of reasoning because Neyman-Pearsion statistical inference is only possible when the data set was given at first.

4. Recommendations for development of Statistical Thinking

On the curriculum for high school level, we would like to propose figure 1 as for the reform framework for any subject of teaching. Here, we finalize the recommendations to develop statistical thinking. It includes traditional statistical thinking based on Neyman Pearson Statistics and problem solving through PPDAC cycle is necessary to critique the approach for the big data, Neyman Pearson Statistics is necessary to learn critical discussion for the data. On the other hands, Data Sciences and Bayesian Statistics are necessary for the handling Big Data on the creative inquiry to produce new findings and ideas such as predicting the future. On figure 1, exploratory data handling includes the five phases of data handling and using of Bayes' Theorem as well as Exploratory Data Analysis. Here, we describe recommendation and considerable points for secondary school level however we do not specify the name of academic subject such as mathematics or informatics because statistical thinking can be developed and applicable for any subjects. We specify the topic to one of three pillars on figure 1 as the most related pillar however please note any subject can be seen from three pillars.

In relation to Neyman Pearson Statistics

Neyman Pearson Statistics is academic subject and existed as the school curriculum content. Formal statistical inference was embedded into the school curriculum at every economy. Law of large number bridges sampling and variability and mathematical probability for part whole relation. Statistical inference such as statistical test under the normal distribution and binominal distributions. Ongoing curriculum reform, informal statistical inference has been enhanced to embed prior exploration of data before teaching formal statistical inference in relation to PPDAC cycle. It enable students to learn the meaning and object/entity of statistical inference before learning formal inference. Because of formal statistical inference is not compulsory the opportunity of informal statistical inference is the opportunity for all students what statistical thinking is on Neyman Pearson Approach. Statistical modeling is also enhanced in relation to visualization with computer as well as the problem solving with PPDAC cycle.

In relation to Data Science

Data Science itself needs higher mathematics and skills for programing which learned at university level. At high school level, students need to experience data handling in relation to five phases: Patterns and relationships from data, Questions, Objectives, Data mining, Understanding and/or designing. Those phases can be learned any subject such as social study and so on. Basic skills to use computer on the web is necessary. Further skills to use scatter diagram is also need to be learned. At the high school level, students also need to learn odds, part-part relationship to develop decision making tree (see chapter 2, exemplar 3&4).

In relation to Bayesian Statistics

Bayesian Statistics is useful for the era of Big Data and AI. Bayes' theorem and the treatment like the exemplar of COVID-19 is necessary to learn at the high school level. Probability function and likelihood need to be introduced. At the same time, the idea of likelihood will be introduced for decision making tree from the middle school as odds which is discussed as part-part relationship instead of mathematical probability as for part-whole relationship.

Necessary Preparation at the middle and primary school levels

For introducing these content at high school level, we also needs to re-consider the middle and primary school curriculum.

The first issue at the middle school levels is visualized modeling for considering the various viability with computer which produces the object of mathematization at the high school. Various visible models are histogram, boxplot, stem-and-leaf display and so on. Students needs to learn them depending on the way of data handling with using computer from the middle school such as how change the mode depending on the range for histogram. Representative value is necessary but usefulness of it can be learned through data handling with computer such as the comparison of variability of distribution with the way how to explain it.

Second issue is introducing the idea of odds and probability from primary school level. Probability has been avoided because it was recognized as the matter of Gambling however the idea of odds (part-part relationship) and probability (part-whole relationship) has been used on our life in the current era. At the upper grade of primary school, students usually study the ratio and rate. Ratio and rate is a key for visual representation of statistics. From the primary level, students need the opportunity to learn the situation of decision making by using rate and ratio with its visible representations such as Band graph.

These are recommendation for the curriculum reform to develop statistical thinking on the big data.

REFERENCES

delMas, R. C. (2002). Statistical literacy, reasoning and learning: A commentary. Journal of Statistics Education, 10(3). Retrieved from http://jse.amstat.org/v10n3/delmas_discussion.html

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. International Journal of Technology Enhanced Learning, 4(5/6), 304–317.

Gal, I. (2005). Adults' Statistical Literacy: Meanings, Components, Responsibilities. International Statistical Review 70(1). 1-51.

González, O., Isoda, M., & Araya, R. (2019). A new framework for statistical thinking in times of big data and digital economy. Manuscript submitted for publication to Statistical Education Research Journal.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. Seoul: McKinsey Global Institute.

Storey, V. C., & Song, I. Y. (2017). Big data technologies and management: What conceptual modeling can do? Data and Knowledge Engineering, 108, 50–67.

Suematsu, M. (to appear). Bayesian Statistical Inference for High School Mathematics by Using the Example of COVID-19. (In submission).

Watson, J. M. (2017). Linking science and statistics: Curriculum expectations in three countries. International Journal of Science and Mathematics Education, 15, 1057–1073.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. International Statistical Review, 67(3), 223–265.