

The role of computational intelligence in developing countries: data access, manipulation and interpretation natural drivers for socio-economic transformation

MWITONDI, Kassim

Available from Sheffield Hallam University Research Archive (SHURA) at:

https://shura.shu.ac.uk/15780/

This document is the Accepted Version [AM]

Citation:

MWITONDI, Kassim (2017). The role of computational intelligence in developing countries: data access, manipulation and interpretation - natural drivers for socioeconomic transformation. In: Computational Intelligence for Societal Development in Developing Countries (CISDIDC), Sheffield Hallam University, 17 February 2017. (Unpublished) [Conference or Workshop Item]

Copyright and re-use policy

See http://shura.shu.ac.uk/information.html

The Role of Computational Intelligence in Developing Countries

Data Access, Manipulation and Interpretation - Natural Drivers for Socio-Economic Transformation

The First Seminar on Computational Intelligence for Societal Development in Developing Countries (CISDIDC): The Role of Computational Intelligence in Developing Countries Friday, 17th February 2017

Kassim S. Mwitondi (PhD)

Sheffield Hallam University; Computing and Communication Research Centre <u>k.mwitondi@shu.ac.uk</u> CC <u>mwitondi@yahoo.com</u>

https://www.shu.ac.uk/about-us/our-people/staff-profiles/kassim-mwitondi

Presentation outline/Sign-posting

□ **PART I: WHAT** this seminar is about – Tackling Challenges

- Semantics: Computational Intelligence and Development
- > Underlying Challenges and Objectives
- □ PART II: WHY? Motivation, justification, current state
 - > Challenges and opportunities
 - > Transforming societies via data, information and knowledge
- □ **PART III: HOW** do we do it?
 - > Frameworks for accessing and sharing data, tools and skills

Interdisciplinary approach; knowledge transfer partnerships

POTENTIAL OUTCOMES: Data sharing frameworks in health, education, agriculture, tourism, international trade, revenue collection, infrastructure, governance & other sectors.
 Concluding remarks, Discussions, Questions and Answers

Logical and Lexical Semantics of the Theme □ The meaning, presupposition, implication and the relationship among the words of the theme... Unlike mathematics or statistics, Computational **Intelligence** is not an established science □ It still has no established foundations as it is a portfolio of tricks and methods (formal and informal) used to address challenges mankind faces. □ Typically, the methods are "**probabilistic**" seeking to deal with uncertainty in the same way as human reasoning - making use of inexact/incomplete data/knowledge in adaptive control environment.

Logical and Lexical Semantics of the Theme

- There has never been a clear distinction between "developing" and "developed" countries. The International Monetary Fund (IMF) distinguishes between "advanced" and "emerging" market economies without any strict criteria.
- Despite labelling 159 countries as DCs, the United Nations (UN) doesn't have an official definition of what a DC is.
- The World Trade Organisation (WTO) lets members define themselves as "developing" or "developed."
- The World Bank previously used a cut-off point the bottom two-thirds of gross national income (GNI) as DCs and from last year, it has removed from its Development Indicators the distinction between the two concepts.
- □ We can see why. **Qatar, Kuwait** and **Singapore** all have a higher per capita income than the **US.**

Computational Intelligence and Developing Countries

- Dean and Kanazawa (1989) provide a natural link of CI to DC in the latter's current context of socio-economic, political and cultural transformation. In particular, the authors describe a model of causal reasoning that accounts for knowledge concerning cause-and-effect relationships.
- Their motivation was based on the premises that "...reasoning about change requires predicting how long a proposition, having become true, will continue to be so." Consequently, in the absence of perfect knowledge, it is natural to expect an agent to be "...constrained to believe that a proposition persists indefinitely simply because there is no way for the agent to infer a contravening proposition with certainty."
- Their foundations of causal reasoning span back to the ancient Greeks and forth into recent advances in computing as it addresses classical problems in temporal reasoning - which fits nicely in the domain of DC.
- □ In linking CI to DC, we assert that *"intelligence controls reasoning"* and decision making-it is therefore a vital attribute we need to *"develop."*

CORRELATION BETWEEN KNOWLEDGE AND DEVELOPMENT

- Our challenges opportunities, strengths and weaknesses lie within the domains of what we know, what we don't, what we can and can't do.
- Other things being equal, we succeed because "we know" and we fail because we lack knowledge and so Alexander the Great and the ancient Greeks had to consult the Oracle at Delphi before going to war.
- □ We still consult fortune tellers, consultants and spiritual leaders before getting married, before making landmark decisions.
- Bioinformatics and computational biology have opened new avenues in biomedical Sciences – they gave us the human genome.
- □ Business Intelligence we believe, is a way we should do business today.
- Criminology utilises new technologies in controlling our borders (not walls) we can track and model emotions; lie detection etc.
- **Food and water are in short supply space/terrestrial weather can help.**
- □ World populations are growing. We track and model migration trends
- Satellite maps of disappearing ice and extreme weather events warn us of the impact of climate change – unless you believe it is a Chinese hoax.

DC's CORE CHALLENGE: TACKLING THE POVERTY CYCLE



- □ An easy to understand GCSE poverty cycle
- □ Knowledge is key to addressing each node
- □ Each node consists of measurable data attributes
- □ These must be harnessed, stored, modelled & shared
- DCs' approach to Big Data must be adaptive
- Computational intelligence to identify and address real challenges on the ground
- DCs need reliable data repositories, shared across disciplines and boundaries
- □ Galileo Galilei: Measure what is measurable and make what is not measurable measurable

Source: https://whsgcsegeographysupport.wordpress.com/the-development-gap/

MULTIVARIATE NATURE OF DEVELOPMENT



- Each of the 17 goals is multivariate and they are highly correlated
- □ In a typical DC, these data attributes are high-dimensional, multi-faceted and fragmented
- □ We can cluster...
- □ We can classify...
- □ We can associate...
- □ We can predict...
- $\Box \quad \text{Or can we?}$
- □ Do we know...
- □ WHO is doing or did
- □ WHAT?
- □ WHERE?
- □ WHEN?
 - HOW?

ADDRESSING GOAL #2 (ZERO HUNGER) THROUGH OPEN DATA



- Global The Open Data for Agriculture & Nutrition initiative 2017) (GODAN, asserts that the the solution to challenge of hunger and malnutrition "Zero Hunger lies within existing, but unavailable, often agriculture and nutrition data".
- Hence, it promotes Open Data policy to make information about agriculture and nutrition available in order to help address the challenge of global food security.

ADDRESSING GOAL #2 (ZERO HUNGER) THROUGH OPEN DATA



- Optimal linking of farming communities with potential markets/customers
- Visualisation of spatiotemporal aspects of local farming, changing patterns
- Reducing wastage by linking with families, schools, Maasais and Sukumas...
- Its scheme, Open Geospatial Science is built upon the premises that "...scientific knowledge...develops more rapidly and productively if openly shared..."
 - Its key ingredients are enshrined in the Principles – "Open Source geospatial software, Open data, Open standards ,Open educational resources, and Open access to research publications.

ADDRESSING GOAL #4 (QUALITY EDUCATION) THROUGH OPEN DATA



- UNESCO-UIS (2016) reports a number of challenges in learning and assessment.
- Most DCs' educational systems either have colonial influences or have been adopted from "developing" countries. Put these together and cross-nationally comparable data becomes difficult/impossible to attain.
 - Prado and Marzal (2013) highlight the importance of data in society by recommending that learning institutions at all levels enhance their contribution to data & information literacy.
- Sets of core competencies and contents that can serve as a framework of reference for inclusion in data and/or information literacy programs.

Hence, the General Message and Questions are...



- Need interdisciplinary learning, sharing data, tools and skills.
- Do what we can do now, with what we have-data, skills, tools
- Some governments call upon people to pray for rain during dry seasons. We do...
- □ Can we record/share data on rainfall?
- □ Can we use them to optimise farming?
- □ Provide clean water?
- Can we teach basic association rules to the youngsters based on these phenomena?

□ YES, WE CAN...

ADDRESSING GOAL #9 (INDUSTRY, INNOVATION & INFRASTRUCTURE)



- Big Data? The digital divide still defines the global imbalance (Mwitondi, 2009).
- DCs must fully utilise existing data flow infrastructure alongside modernisation programmes.
- □ Forming adaptive research synergies between the academia and the private (KTP, more later).
- What have the DCs got? The Japanese have very little in the form of land and resources but plenty in creativity & innovation.
- Industrialisation derives from STI, it transforms agriculture, transport, utility supply, markets (home & abroad), mind sets and, above all, it promotes brand names – as did Galaxy/iPhone...

ADDRESSING GOAL #9 (INDUSTRY, INNOVATION & INFRASTRUCTURE)



- Reforms in legislation, technology, infrastructure and mind sets – crucial human factors come in the form of "education" and "good governance".
- These associations must be uncovered and used as inputs into development strategies.
- □ If overlooked, the industrialisation equation becomes unsolvable.
 - Everything that we do in life as individuals, families, ministries or countries we set off from the simplest to the complex.
- □ Institutional arrangement among various bodies is inevitable.
 - Sharing what is known, what is accessible helps consolidate whatever strategies there may be.

WHAT IS AVAILABLE AND ACCESSIBLE?

High-speed fibre optic connectivity has provided robust Internet backbone and hence general ICT development across the world.



- □ Africa One: It was perceived as a great idea at the turn of the last millennium.
- The project idea never took off due to financial, political and technical hurdles
- The hurdles could probably have been overcome, had there been "useful" data sources and potential flows across the continent at the time.

WHAT HAVE THE DCs ACHIEVED? AND WHAT ELSE CAN BE DONE?

- □ Twenty years on, a number of submarine and terrestrial fibre optic cables connect African countries to the Internet and more are planned for major cities!
- □ They are expected to revolutionise ICT trends around the continent and enhance its socio-economic & cultural transformation. In major cities?
- □ Beneficiaries from these developments have mainly been telecoms companies users benefitting via telephone banking, e-Health and educational schemes.
- □ Could and should we do more with respect to the development goals? Surely.



Source: https://mybroadband.co.za/news/broadband/106353-here-is-who-controls-the-internet-in-south-africa.html

Many Challenges and Opportunities Still Lie Ahead

- □ Many parts of the DCs are yet to significantly address existing challenges and potential opportunities. In Africa, there are many who believe that potential benefits are constrained by its spatial coverage and so satellite technology is the best way to reach rural areas.
- □ Unfortunately, both options still aren't standard in most DCs including Kenya a country viewed as a leader in African technology innovation.



Source: http://www.bbc.co.uk/news/business-32079649

- The Isizwe Project in South Africa focuses on wireless technologies to bring free internet to everyone in South Africa by installing wi-fi hotspots in low-income areas.
- Isizwe considers satellite and microwave technologies as the future, with Wi-Fi, 3G and 4G providing extensions to small businesses and homes.
- This assertion is probably supported by the huge success of mobile banking across many parts of the developing world.
- Key prerequisites are policies GVT commitment), data, skills, tools, design and management.

PUTTING KNOWLEDGE AND DEVELOPMENT INTO PERSPECTIVE

- Chen and Dahlman (2004) assessed the effects of knowledge on economic growth of 92 countries over forty years (1960-2000) and concluded that human capital, domestic innovation, technological adaptation, and the level ICT infrastructure had a significant effects on long-term economic growth.
- Globally, ICT R&D is geared towards productivity, innovation and profitability – hence, proprietary in nature. In the DCs ICT education produces "systems supporters" who are often detached from mainstream socio-economic phenomena.
- □ This trend creates gaps between the elite & the rest of the population; between the academia & industry. It forms knowledge clusters within these domains and so restricts access to data, knowledge, tools and skills. Hence, as we talk of CI and DC, coupling and cohesion of these clusters are fundamental.

JUST HOW CAN THE GAPS BE NARROWED?

With all the hype about Big Data and ICT, the digital divide still defines socio-economic imbalances between the two hemispheres (Mwitondi, 2009). Bridging this gap is an uphill and multidimensional task that requires co-ordinated interventions.

□ GOVERNMENTS MUST TAKE THE LEAD

- Weathington (2017) underlines the need for CEOs in companies to lead Big Data initiatives in organisations – as part of corporate strategy.
- > DC Governments must play key roles in promoting open source as part of their "development strategy" otherwise they are doomed to fail.
- > A top-down approach is recommended. The political elites at the very top, must champion these initiatives. It seems to be working in Rwanda.

KNOWLEDGE TRANSFER PARTNERSHIPS

- To bridge this gap, DCs must optimally utilise existing data flow potentials
 novel and conventional and the best way to achieve this is by
- > ...forming adaptive research synergies between the academia, private sector and other public institutions. These initiatives need funding.
- > <u>Use business-political synergies to enforce contractual clauses...</u>

How? Collectively Set Vision, Values And Measurable Objectives (VVMO)

- □ Intersecting socio-economic, cultural and environmental development is central to development.
- Central & local governments, NGOs and individuals must come together with shared VVMO. These may vary but they must be adapted along the way
- □ How we "shape" our development spectrum is extremely central to attaining VVMO.
- □ Our VVMO require invariance this is a challenge.
- □ Geometry describes "shape" as what remains after removing the effects of "translation and rotation" that is, clearing perturbations, crises or disagreements.
- Coon (1946) describes universality of human societies from various perspectives. Development is "invariant" and data on attributes that describe it are crucial.

SOME OF THE STEPS ALREADY TAKEN



Given limited resources - financial, treated nets, rapid malaria tests, a malaria control project team in Zambia (a couple of years ago) resorted to data visualisation to decide spray targets. But with data from health centres coming from as far away as 25km by

traditional means of transport, analyses were often of little use.

The Visualize No Malaria (http://visualizenomalaria.org/) partnership between the Zambian Government and the private sector now provide the team with access to powerful and timely analyses, or does it?

Source: http://visualizenomalaria.org

- □ The project "boosts economies" by cutting malaria treatment costs (currently over US\$12bn a year)
- □ It frees up scarce public health resources (40% is spent on malaria)
- □ It strengthens education (50% of absenteeism due to malaria).
- □ <u>Its visualisation tools are loaded...</u>
- Most DCs run similar projects funded differently and quite often not talking to each other.
- Data and/or information gained often get lost between projects, ministries, funding arrangements.
- Bill and Melinda Gates Foundation, the WB, USAID, DFID, JICA etc
- □ We should retain data to help describe our universality with respect to "development").

SOME OF THE STEPS ALREADY TAKEN



□ Fancy names to attract admissions? May be... □ What are their outcomes? □ Most of start at Uni-level □ How wide-spread? □ How accessible? □ Do they help us curb real life challenges? Yes... □ Could do more? Yes... □ What is wrong?

The universality of global phenomena stipulates that our strategies for tackling real-world challenges start early <u>http://www.bbc.co.uk/news/education-38938519</u>

STEPS THAT NEED TO BE TAKEN

Interdisciplinary approach to problem solving must be adopted from much earlier in professional careers than the current practice is.
 Merging disparate disciplines based of their

analytical requirements, educational systems can build interdisciplinary bridges across curricula.

Statistics for biologists

Home Practical guides | Statistics in biology | Points of Significance | Other resources



There is no disputing the importance of statistical analysis in biological research, but too often it is considered only after an experiment is completed, when it may be too late.

This collection highlights important statistical issues that biologists should be aware of and provides practical advice to help them improve the rigor of their work.

Nature Methods' Points of Significance column on statistics explains many key statistical and experimental design concepts. Other resources include an online plotting tool and links to statistics guides from other publishers.

Image Credit: Erin DeWalt

Unifying concepts in various disciplines is not novel to research communities. Mwitondi (2003) discusses the link between statistics & computer science. From similar premises we can define categories of data literacy, competencies and resources needed.

Examples of Composite Challenges in Southern Tanzania

- High Poverty in areas surrounding National Parks
- 2. Limited local benefits
- 3. Poor infra/high travel costs
- Increase in concentration of economic activity in the ecosystem:
 - (a) Agricultural expansion
 (b) Pressures on carrying capacity
 (c) Deforestation
- 5. Poaching
- 6. Illegal logging



Further Examples of Composite Challenges in Southern Tanzania

- About 56% of flow to the Mtera and Kidatu power plants – make 50% of total TZ hydropower capacity
- Rapidly Expanding Irrigation in and around the Usangu Plains (BRN)
- Reduced Environmental Flows, affects Mtera and Kidatu
 - Tourism significance Ruaha NP is the core of the southern circuit but without water no wildlife/tourism
 - Downstream economic activities Poverty impacts; levels of water use/stress puts livelihoods at risk
 - Human-wildlife conflicts expanding during, especially in dry seasons
 - Settlements without schools





4/15/2017

Measuring What is Measurable



- A Jamaican-born lady named <u>Sue Stolberger</u> has been regularly measuring the depth of the Grand Ruaha & compiling various fact files for over 12 years.
- Sue is not a research scientist. She is inspired by her love for art and nature.
- Yet, she maintains thrilling rudimentary data repository.
- □ How about empowering local schools to carry out this exercise and share data?

On-going initiatives are supported by the central and local GVTs, WB and others

SUPPORTING OPEN SOURCE DC GOVERNMENTS MUST PLAY KEY ROLES

- \Box Proprietary tools have thrived for a reason R&D investment.
- Open software development is prone to more risks than proprietary open testing exposes it to hackers; its vulnerability is compounded by the common phenomenon in programming - component dependency.
- Open source vulnerability management is and will remain a key concern. With a variety of contributors from around the world, development management teams may find themselves pushed to the edge of propriety.
- □ NASA code (<u>https://code.nasa.gov</u>) and
- □ R (<u>https://www.r-project.org/about.html</u>) (<u>http://developer.r-project.org/</u>) have stringent measures in place and they are well-funded.
- □ Weathington (2017) underlines the need for CEOs in companies to lead Big Data initiatives in organisations as part of corporate strategy.
- DC Governments must play key roles in promoting open source as part of their "development strategy" otherwise they are doomed to fail.
- □ A top-down approach is recommended. The political elites at the very top, must champion these initiatives. At least it seems to be working in Rwanda.

SUPPORTING OPEN SOURCE APPLICATIONS ADAPTIVE EDUCATIONAL & VOCATIONAL PROGRAMMES NEEDED

- □ R&D funding in most DCs is predominantly foreign-dependent.
- NASA Europa Challenge OpenCitySmart for developing tools for city-related infrastructure management is based on Europe's INSPIRE Directive to guide project development.
- □ While data management tools may be universal, traffic control in Dar, Lagos or Mumbai may need innovative approaches...
- ...innovations that don't necessarily derive existing sophisticated traffic control systems you see in metropolitan London, Tokyo or New York, but, possibly portable systems of smart apps focusing on "what works". Examples already exist
- □ The same approach should be adopted in tackling each of the 17 goals. But remember, they are inseparable and so they require a coherent, unified approach requiring your CI skills and tools.

Typical Examples of Collaborative Setting of VMOs

DATA

STRU

SHARING DATA, RESEARCH FINDINGS, SKILLS & TOOLS



COMPARATIVE STUDIES

Enabling comparative studies as a way of positively contributing towards understanding the clinical and/or pathogenetic features of diseases like

- Sickle Cell Anaemia (SCA)
- Systemic Lupus Erythematosus
- ≻ Various types of cancer
- >Lymphatic Filariasis
- ≻Malaria
- ≻HIV/AIDS

Collect/manage data for modelling disease patterns.

Investigation of race-specific diseases.
Developing a tracking index for various diseases that could be used as a benchmark for intervention programmes.

SUPPORTING OPEN SOURCE APPLICATIONS ADAPTIVE EDUCATIONAL & VOCATIONAL PROGRAMMES NEEDED

- □ R&D funding in most DCs is foreign-funded and often owned.
- NASA Europa Challenge OpenCitySmart you can join the initiative. But note that it is inspired by Europe's INSPIRE Directive and so it may not be directly transferable.
- Data management tools may be universal, but traffic control in Dar, Lagos or New Delhi may need adaptive solutions.
- You should not seek to emulate or improve upon the sophisticated traffic control systems you see in London, Tokyo or New York, but focus on "what works", while you remain "smart". These are likely to be portable systems like the mobile apps, currently in use in major African cities.
- □ Adopt the same philosophy in tackling each of the 17 goals. They are correlated, hence require coherent, unified approach.
- Dynamics of the Grand Ruaha (km 475 long over 83,970 square kilometres) can be tracked and monitored without the need for satellites. You just need to be proactive, innovative and "open".

BEARING IN MIND THE FOREGOING CONSIDERATIONS, YOU CAN THEN APPLY YOUR SKILLS AND ALL THE NOVEL TOOLS YOU HAVE DEVELOPED



BEARING IN MIND THE FOREGOING CONSIDERATIONS, YOU CAN THEN APPLY YOUR SKILLS AND ALL THE NOVEL TOOLS YOU HAVE DEVELOPED

$$P(Y|X_1, X_2, ..., X_{\lambda}) = \frac{P(X_{\lambda}|Y)P(Y|X_1, X_2, ..., X_{\lambda-1})}{\int P(X_{\lambda}|Y)P(Y|X_1, X_2, ..., X_{\lambda-1})dY}$$

ALLOCATION RULE ERRORS DUE TO DATA RANDOMNESS

POPULATION	TRAINING	CROSS VALIDATION	TEST
ψ _{D,POP}	$\psi_{D,TRN}$	ψ _{D,CVD}	$\psi_{D,TST}$
	C		

Source: Mwitondi (2003)

$$\Psi_{D,CVD} = \sum_{k=1}^{K} \sum_{i=1}^{N} \pi_k P(X_i \in C_k | Y \notin C_k)$$

References and Bibliography

- 1) Chen, D. and Dahlman, J. (2004). Knowledge and Development: A Cross-Section Approach. Policy Research Working Paper;No.3366. World Bank, Washington, D.C.
- 2) Coon, C. (1946). The Universality of Natural Groupings in Human Societies; The Journal of Educational Sociology, Vol. 20, No. 3, pp. 163-168, American Sociological Association.
- 3) Dean, T. and Kanazawa, K. (1989). A model for reasoning about persistence and causation Computational Intelligence; Vol 5, Issue 2, pp 142–150.
- 4) Dreher, A., Gaston, N. and Martens, P. (2008). Measuring Globalisation: Gauging Its Consequences; Springer, ISBN-13: 978-0387740676.
- 5) GODAN (2017). Global Open Data Initiative; <u>http://www.godan.info/about</u>
- 6) Kuznetsov, Y. (2006). Diaspora Networks and the International Migration of Skills: How Countries Can Draw on Their Talent Abroad; WB Publications, ISBN-13: 978-0821366479
- 7) Mwitondi, K. S. (2009); Tracking the Potential, Development, and Impact of Information and Communication Technologies in Sub-Saharan Africa; A book chapter in: Science, Technology, and Innovation for Socio-economic Development: Success Stories from Africa; International Council for Science (ICSU-ROA), ISBN 978-0-620-45741-5.
- 8) Prado, J. and Marzal, M. (2013). Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents; Vol. 63, Issue 2, pp. 123–134, ISSN 0024-2667
- 9) UNESCO-UIS (2016). Laying The Foundation To Measure Sustainable Development Goal 4; Sustainable Development Data Digest, ISBN 978-92-9189-197-9.
- 10) Weathington, J. (2017). Why CEOs Must Lead Big Data Initiatives; <u>http://www.techrepublic.com/article/why-ceos-must-lead-big-data-initiatives/</u>

ANY QUESTIONS?