Information poverty - Connecting every child to opportunity

Naroa Zurutuza[1], Vedran Sekara[1], Anna Bethke[2], Christopher Fabian[1] and Manuel Garcia-Herranz[1]

[1]UNICEF Office of Innovation
[2]Intel AI Lab
Corresponding Authors: nzurutuza@unicef.org and mgarciaherranz@unicef.org

Abstract

For a child to be nutritionally healthy, 1,500 calories are required each day. Can similar measures be defined with respect to information? If so, how many kilobytes and of what type does a child need to be "information healthy" and have equitable access to opportunity and choice? UNICEF is creating an open source platform to measure a child's access to information and use this to build sustainable infrastructures and programmes that connect every child to opportunity and employment [1]. Here we explore ways Artificial Intelligence can help to connect every child in the world to the information they require to prepare them for the future.

Being information poor

The 21st century has seen a world transformed by connectivity. At a global scale connectivity has reshaped the lives of many in what has been called the information revolution. News, social connections, jobs, education, health services, and opportunities and choice in general have all adapted, and continue to change, to leverage the power and possibilities of this new age. This change, however, comes with challenges. Some of them have been extensively studied like issues of information overload [2], information bubbles [3], and the spread of fake news [4], while other, equally important topics, have been insufficiently studied. Underlying all of the Sustainable Development Goals (SDG), as well, is a thread of information and opportunity [5]. We will not be able to fully achieve these ambitious benchmarks for change if certain populations, particularly the young, are disconnected or ill-prepared for the future.

A way of describing this sub-strata of the SDGs is through the lens of information poverty. Information is a driving force of opportunity and choice [6]. What does it mean to be information poor? This question is a highly complex one encompassing multiple dimensions of deprivation. Information poverty can limit many types of opportunity, from access to job skills, financing, creative and social networks, and the ability to develop the skills necessary for global citizenship [7]. Due to the complexity of this problem new sources of data about both information access and young peoples' needs, along with AI and machine learning are required to help us understand information and infrastructure gaps of the most vulnerable. UNICEF, and partners, can use new models and data-driven approaches to work with governments to ensure provision of, and access to, relevant and necessary global digital public goods.

Mapping the physical landscape

In order to connect young people to opportunity and employment, we first need to understand where they live and what the infrastructure gaps around them are: Do children have access to schools and

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

education? Are they part of the information society? To answer these questions, UNICEF is mapping every school in the world and measuring their connectivity in real-time [8].

AI has been used to identify patterns in satellite imagery [9–11]. Nevertheless, schools, especially in rural areas of low and middle income countries present an additional challenge, since existing pre-trained neural networks such as ImageNet or ResNet [12, 13] are biased towards imagery and landscapes for high income regions. Schools are central for UNICEF's work as they act as hubs within communities. Thus, UNICEF is working with governments to create a comprehensive school data corpus that accounts for regions that are not traditionally accounted. UNICEF is then using this data corpus to work with partners like UC San Diego and Development Seed to explore the use of high resolution satellite imagery and Deep Learning techniques to detect schools [14]. The initial results for the country of Liberia show great promise (76% overall accuracy) and demonstrate the feasibility of using AI to detect schools. This is, however, just a pilot project, and considerable work is required to put this into practice to be able to detect schools globally across a variation of contexts (urban, rural, regional). Similarly, remote sensing and deep learning techniques have shown promise in mapping infrastructures, including: roads [15], water bodies [16], urban/suburban structures [17] and power lines [18]. Can other sources of data such as mobility patterns, social media posts, or other data be used to fill in our knowledge gaps and if so how can we further the use of AI in this task?

Having the base layer of school location and school connectivity allows us to start addressing infrastructure gaps and bring connectivity, information and opportunity to disconnected communities. However, how do we measure, at scale, whether schools are connected to the internet? Can AI help us do this and optimize its delivery?

Mapping the digital landscape

Complementary to addressing infrastructure gaps and connecting every school to the Internet, we need to make sure that children have access to the right content (i.e. in their own language, locally produced, and relevant to their context). Previous work has shown a correlation between the development of network infrastructure and the creation of local content [19]. Yet, the distribution of information and content across different geographies and languages still remains uneven [20]. While Wikipedia is a shining beacon of information it is a glaring example of how information is segregated, with different language editions varying dramatically in how comprehensive they are [21]. As a result, a majority of languages contain only a small fraction of the sum of information that exists for the English version. Recent advances in AI and Natural Language Processing (NLP) can aid us in solving and detecting these information gaps by analyzing and measuring content availability.

Moreover, not all content is equal. As such, we value content that deals with local community higher than yellow press content, for example. NLP techniques can aid us in distilling and distinguishing content that is relevant to a specific context and allow us to develop mechanisms that can quantify content gaps. Finally, we can build human-in-the-loop solutions that combine the power of AI and human expertise to fill these gaps.

Opportunities

Information and communication technologies have changed the world, creating new opportunities for many but also deepening inequalities for others. With training sets and data from mobile network operators, satellite imagery providers, ministries of education and finance, and more, UNICEF can be a unique interlocutor between the needs of a billion young people and the opportunities presented by emerging technology.

In a world that is increasingly more fragmented, and where the difference between having a smartphone and a basic "dumb phone" can mean entirely different levels of access to education, health, and other opportunities it is neither fair nor practical to create a population that is starved of vital information. Being able to apply ML and AI to the new types of data generated by UNICEF and partners can allow us to look at the intersection of infrastructure needs, learning materials, and job opportunities for young people - and potentially begin to optimize some of these to create a more fair and equitable world.

References

[1] MagicBox – Information Poverty. *UNICEF Innovation Builds and Scales Innovations That Improve Children's Lives around the World* (2018). `https://unicefstories.org/magicbox/infopoverty/`.

[2] D. Bawden, and L. Robinson. *The dark side of information: overload, anxiety and other paradoxes and pathologies.* Journal of information science, 35(2), 180-191 (2009).

[3] E. Bakshy, S. Messing and LA. Adamic. *Exposure to ideologically diverse news and opinion on Facebook.* Science, 348(6239), 1130-1132 (2015).

[4] DM. Lazer, et al. *The science of fake news.* Science, 359(6380), 1094-1096 (2018).

[5] United Nations General Assembly. *Transforming our world: the 2030 Agenda for Sustainable Development* (2015).

[6] LM. Sene. *Connectivity as engine for productivity among smallholder peanut farmers in Senegal.* In 2015 Conference, August 9-14, 2015, Milan, Italy (No. 212263). International Association of Agricultural Economists (2015).

[7] United Nations Children's Fund. *State of the World's Children 2017: Children in the digital world*. UNICEF, New York (2017). `https://www.unicef.org/publications/index_101992.html`.

[8] *Project Connect*. `https://projectconnect.world/`.

[9] Y. Chen, et al. *Deep learning-based classification of hyperspectral data.* IEEE Journal of Selected topics in applied earth observations and remote sensing 7.6 (2014): 2094-2107.

[10] M. Vakalopoulou, et al. *Building detection in very high resolution multispectral data with deep learning features*. Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International (2015).

[11] Q. Zou, et al. *Deep Learning Based Feature Selection for Remote Sensing Scene Classification*. IEEE Geosci. Remote Sensing Lett. 12.11 (2015): 2321-2325.

[12] J. Deng, et al. *Imagenet: A large-scale hierarchical image databasex*. Computer Vision and Pattern Recognition, 2009. IEEE Conference on. Ieee (2009).

[13] K. He,et al. *Deep residual learning for image recognition.* Proceedings of the IEEE conference on computer vision and pattern recognition (2016).

[14] GBDX for Sustainability Challenge, (2017). `https://blog.digitalglobe.com/industry/gbdx-for-sustainability-challenge-mapping-every-school-in-the-world-and-reducing-the-digital-divide-in-education/`.

[15] M. Mokhtarzade and MJV. Zoej. *Road detection from high-resolution satellite images using artificial neural networks.* International journal of applied earth observation and geoinformation 9.1 (2007): 32-40.

[16] PS. Frazier and JP. Kenneth. *Water body detection and delineation with Landsat TM data.* Photogrammetric engineering and remote sensing 66.12 (2000): 1461-1468.

[17] JR. Jensen and DC. Cowen. *Remote sensing of urban/suburban infrastructure and socio-economic attributes.* Photogrammetric engineering and remote sensing 65 (1999): 611-622.

[18] R. Bhola, et al. *Detection of the power lines in UAV remote sensed images using spectral-spatial methods.* Journal of environmental management 206 (2018): 1233-1242.

[19] Organisation for Economic Co-operation and Development. *The relationship between local content, internet development and access prices*. OECD (2013). `https://www.oecd.org/internet/ieconomy/50305352.pdf`.

[20] S. Ojanpera, et al. *Engagement in the knowledge economy: Regional patterns of content creation with a focus on Sub-Saharan Africa.* (2017): 33-51.

[21] E. Wulczyn, et al. *Growing wikipedia across languages via recommendation.* Proceedings of the 25th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, (2016).