Japanese Journal of Radiology (2019) 37:351–353
https://doi.org/10.1007/s11604-019-00833-3

SPECIAL REPORT

The day when computers read between lines

Kei Yamada[1] · Susumu Mori[2]

Received: 6 March 2019 / Accepted: 20 March 2019 / Published online: 25 March 2019

Abstract

There is a growing notion that artificial general intelligence (AGI) will replace some of the work done by trained professionals, including physicians. This idea, however, seems to have logical leap; herein, we discuss three problems that are significant barriers to this. First, the ground truth is difficult to provide in the majority of medical conditions. Second, the electronic medical record (EMR) only covers a portion of the information that is crucial for patient care. This makes the data in the EMR a suboptimum material for creation of AGI. Third, there are decision-making processes that cannot be captured in a way that computers can digest; portions of our thoughts, perceptions, intuitions, and inspirations cannot be translated into numbers or words.

Keywords Artificial intelligence · Artificial general intelligence · MRI · Electronic medical record · Computer-aided diagnosis

Introduction

Enthusiasm for the application of artificial intelligence (AI) in medical practice has quickly grown over the past few years. There have already been a considerable number of exciting results from around the world, including Japan [1–4]. Anticipation of the clinical implementation of AI arises in part from the substantial shortage of radiologists in some countries [5, 6]. Also, there are apparently aspects of the job that radiologists do not enjoy. For instance, radiologists tend to dislike measuring lesion sizes on follow-up examinations. Many radiologists are hoping that these mundane tasks will be soon taken over by computer-aided diagnosis (CAD) incorporated into the picture archiving and communication system (PACS).

There are in fact some AI tools available in the market already, but these are only able to cover limited portions of medical practice. In the long run, however, there may be a day when further advances in technology will lead to the formation of artificial general intelligence (AGI), which has a substantially wider capacity to support our practice. This, in turn, begs the question of whether this will eliminate the need for doctors. Contrary to popular belief that this will be the case [7], we believe that complete automation of the radiologist's work is a long way off. Herein, we discuss three kinds of general problems that make it unlikely to happen any time soon. The first two problems are medical and other is computer science side related. We start with the problems on the medical side.

Difficulty in providing the ground truth

It is well known that computers need "ground truth" for learning. Providing the ground truth often involves "annotation," or labeling data. Some of these annotations are straightforward while others quite difficult. The level of challenge depends on the degree of uncertainty or complexity when making annotation decisions. Below we will try explaining the differences in the degree of challenge by dividing them into five levels of difficulty.

Kei Yamada
kyamada@koto.kpu-m.ac.jp

[1] Department of Radiology, Graduate School of Medical Science, Kyoto Prefectural University of Medicine, Kajii-cyo, Kawaramachi Hirokoji Agaru, Kamigyo-ku, Kyoto City, Kyoto 602-8566, Japan

[2] Kennedy Krieger Institute, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Level 0: single objective variable not requiring further human decision making

In some conditions, the ground truth can be provided by simple one-to-one correlation with existing data, even without further human decision making. For instance, if one intends to build CAD software that discerns the sex of a patient depicted on a given plain film examination, the task is quite simple. One has to only provide the computer with the known sex of the patient together with the images. Noise reduction of MRI is another example. One has only to feed the computer with two sets of MR images, the first set being the images with a single excitation and the second being images with multiple averaging, and the latter one will be used as the ground truth. We define these tasks as being level 0, in which there is no human decision making.

Level 1: medical decision necessary but by a single discipline

The next level of difficulty is those tasks that require medical decision to provide the ground truth, but by a single discipline (e.g., radiology). One such example is the pediatric bone age challenge hosted by the Radiological Society of North America (RSNA) in 2017 [8]. In this competition, bone age, as determined by the expert pediatric radiologists, were used as the ground truth. A few more examples could be detection of brain metastases and segmentations of organs, which require input from radiologists. We define these tasks as being at level 1.

Level 2: medical decision necessary by multiple disciplines

Many diseases have diagnostic criteria. The diagnoses are typically based on decisions made by multiple disciplines. This potentially makes the degree of uncertainty larger than for the aforementioned level 1 tasks. For instance, the diagnosis of multiple sclerosis (MS) may rest on the McDonald criteria, with the diagnosis being based not only on the clinical course of the patient, but also on the MR imaging findings and tests of cerebrospinal fluid (CSF).

Level 3: ground truth with known uncertainty

While most diagnostic tests are designed to be reasonably reliable, some tests have a substantial level of uncertainty. Pathological examination of gliomas is one such example, and there are two major sources for ambiguity. First, it is well known that sampling error is an inevitable issue, and second there is significant interobserver variability among pathologists [9]. Another example is Alzheimer disease (AD), as the clinical diagnoses of AD often changes after autopsy.

Level 4: controversial clinical diagnoses

Certain conditions do not yet have an established consensus as to how one should reach a final diagnosis. For instance, four competing sets of diagnostic criteria can be used for diagnosis of vascular dementia, including Alzheimer's Disease Diagnostic and Treatment Centers (ADDTC), Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV), International Classification of Diseases, 10th revision (ICD-10), or National Institute of Neurological Disorders and Stroke-Association Internationale pour la Recherche et l'Enseignement en Neurosciences (NINDS-AIREN). Some of the other psychiatric disorders have even less consensus. These diseases will not be the ideal use case for AI.

These levels of difficulty are an arbitrary classification and not entirely scientific, but the aim of this categorization is to highlight the fact that providing the ground truth is not always a simple task. In fact, most of the medical conditions will fall into levels 2–4, making annotation a challenge.

Limitations of the EMR

Many of the recent achievements using AI, e.g., speech recognition, self-driving cars, and image recognition, are based on deep learning with big-data. In the medical field, we also have a substantial repository of big data in our electronic medical record (EMR). Some scientists believe that once it becomes possible to feed computers with the big-data of the EMR, a robotic doctor (AGI) will soon follow.

The EMR contains data encoded in the form of numerical data, text, or images. The easiest for computers to digest is numerical data, which includes vital signs and laboratory data from samples such as blood or urine. Text data are now becoming another important source of big-data, thanks to the advent of natural language processing (NLP). This area has become a target of active research, such as text-based risk prediction for patients. Last but not least, the imaging data, including that from radiology, is another vital piece of the medical information puzzle.

If it becomes possible to feed all three forms of data into computers in the near future, will it truly lead to the creation of human-level AI? This may sound like a plausible hypothesis, but we believe that it will not be that simple. This is because the EMR cannot be a comprehensive record of medical practice.

Consider that we enter notes in the EMR before and after procedures. We dictate radiological reports day-in-day-out.

One question we should ask ourselves is whether we record everything that led to our final medical decisions. The clear answer to this is "no." We only record a small percentage of our thoughts. We are gifted with skills to summarize things, and we usually choose to write down only those issues that are pertinent to patient management. We record only the minimum amount of text to justify the choice we make, or just enough to precisely convey our ideas to our colleagues. The reason for spending a minimum amount of time and energy for "recording" is quite obvious: it is not one of our main tasks. Our main tasks are to solve medical problems, decide the next move, and help our patients feel better.

But let us say, in a hypothetical sense, that we could create an institution where the doctors are trained to write down (or dictate) pretty much everything they think. Would this really help us get closer to a perfect big-data for building an AGI?

Not all data can be input into computers

There are certain things that cannot be translated into numbers or words. Experienced doctors tend to use intuition or "gut-feeling" for diagnoses. This apparently also applies to radiologists. When a radiologist is given an unknown case which does not resemble anything that he/she has experienced in the past, they still are often able to offer a reasonable differential diagnoses, and oftentimes, the answer is correct. The implicit processes they experience during these kinds of interpretation processes should be extremely difficult to translate into words.

Another type of information that is difficult to record is nonverbal communication. It is well known that this form of communication constitutes about 70% of our daily exchange of information [10]. Gestures, tone of voice, and facial expressions are some of the most important factors that influence the meaning of our speech. When a radiologist is asked to give the differential diagnoses during a case conference, he/she can put emphasis on those that they feel confident about and de-emphasize the rest, even without directly saying so. Are we able to precisely record this so that the computers can understand the messages we express through body language?

The above-mentioned issues can be considered to be the information at a "higher dimension," which the binary number system cannot be used to record. We would probably need a novel encoding technique to allow for this to happen. For example, emoji may be a partial solution to convey some nonverbal communication, but its use is very much limited to certain facial expressions and gestures. Recording the intuition of doctors will be a much more difficult task. Unless we are able to discover a way of directly encoding our thoughts, there will be no way that we can convey to computers one of the most important aspects of human intelligence. A gigantic leap in technology will be needed to fill in these gaps. Will the new quantum computers solve this problem? I have a "gut-feeling" that this will not be the case. Some computational cognitive scientists believe that it could be a "reverse engineering" of the nervous tissue that will lead to a breakthrough. But this future is yet to come.

In summary, there are three major hurdles in creating comprehensive robotic doctors. First, most diseases we deal with do not have an absolute "ground truth." Second, we currently cannot record everything in the EMR because of time constraints, lack of man power, and impracticality. Last, there is not yet a perfect way to capture our perceptions, thoughts and intuitions. It will take a big breakthrough to record this kind of information. We believe, however, that there will be one day in the future when computers become capable of reading between the lines, and that will be the day when they start catching up with humans.

Compliance with ethical standards

Conflict of interest The authors have no conflict of interest with regard to the content of this manuscript.

Ethical standards The study does not deal with human/animal specimen and is compliant with ethical standards.

References

1. Kobayashi Y, Ishibashi M. Kobayashi How will "democratization of artificial intelligence" change the future of radiologists? Jpn J Radiol. 2019;37:9–14.
2. Ueda D, Shimazaki A, Miki Y. Technical and clinical overview of deep learning in radiology. Jpn J Radiol. 2019;37:15–33.
3. Sakai K, Yamada K. Machine learning studies on major brain diseases: 5-year trends of 2014–2018. Jpn J Radiol. 2019;37:34–72.
4. Higaki T, Nakamura Y, Tatsugami F, Nakaura T, Awai K. Improvement of image quality at CT and MRI using deep learning. Jpn J Radiol. 2019;37:73–80.
5. Nakajima Y, Yamada K, Imamura K, Kobayashi K. Radiologist supply and workload: international comparison: Working Group of Japanese College of Radiology. Radiat Med. 2008;8:455–65.
6. Kumamaru KK, Murayama S, Yamashita Y, et al. Appropriate imaging utilization in Japan: a survey of accredited radiology training hospitals. Jpn J Radiol. 2017;35:648–54.
7. Hinton G. On radiology (2016). https://www.youtube.com/watch?v=2HMPRXstSvQ. Accessed 25 Mar 2019.
8. Halabi SS, Prevedello LM, Kalpathy-Cramer J, Mamonov AB, Bilbily A, et al. The RSNA pediatric bone age machine learning challenge. Radiology. 2018 Nov 27:180736. https://doi.org/10.1148/radiol.2018180736. (Epub ahead of print).
9. Prayson RA, Agamanolis DP, Cohen ML, et al. Interobserver reproducibility among neuropathologists and surgical pathologists in fibrillary astrocytoma grading. J Neurol Sci. 2000;175:33–9.
10. Birdwhistell R. Kinesics and context: essays on body motion communication. Philadelphia: University of Pennsylvania Press; 1970.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.