

How to Cope with Bias in Wellbeing AI? – Towards Fairness in Wellbeing AI by Personal and Long-term Evaluation

Keiki Takadama

The University of Electro-Communications
keiki@inf.uec.ac.jp

Abstract

This paper focuses on the fairness in ML (machine learning) meaning that output of ML should not be “biased” and aims at clarifying bias in wellbeing AI. From the analysis of bias from the viewpoint of healthcare, the bias in wellbeing AI can be reduced by employing the personal and long-term evaluation while many biases in ML arise. To investigate an effectiveness of the personal and long-term evaluation, our previous research conducted the human subject experiment by focusing on sleep of aged persons in care house and found that our wellbeing AI based on the personal and long-term evaluation succeed to extract knowledge for good sleep and to estimate mind change of an aged person from her sleep quality change.

Introduction

“Can ML (machine learning) provide a fair decision?”. To answer this question, this paper starts to explain one example. As you may know, Amazon developed the ML personnel recruitment system but stopped it in 2018 because women do not tend to be recruited in comparison with men due to the reason why most of input data for ML is men’s data (Dastin, 2018). This example suggests an importance of fairness in ML. In other words, the output by ML should be fair or should not be “biased”. What should be noted here is that many healthcare systems based on ML (hereafter we call it as “wellbeing AI”) have also a risk of providing the biased outputs and such outputs are very critical for our daily life. From this fact, this paper aims at investigating what kinds of biases arises in wellbeing AI and how such biases can be reduced to cope with them. For this issue, this paper starts to explain bias in general by focusing on the bias on the Internet and the bias in ML, and then clarifies bias in Wellbeing AI.

Bias on the Internet

According to Baeza-Yeate, the bias on the web on the Internet is categorized as follows (Baeza-Yates 2018): (1) activity

bias, (2) data bias, (3) sampling bias, (4) algorithmic bias, (5) interaction bias, (6) self-selection bias, and (7) second-order bias. The essential difference among them is summarized as follows.

(1) Activity bias

This bias arises from the different number of active/silent users. For example, only top 4% of Amazon users posted the reviews, which means that we cannot receive messages from all users, *i.e.*, they are the biased messages.

(2) Data bias

This bias arises from the different number of data. For example, the number of Westerner face pictures tends to be larger than that of Asian in the face pictures in dataset such as MS (Microsoft) celebrity dataset.

(3) Sampling bias

This bias arises from the fact where the sampled data is not always followed by true distribution. For example, an asthma patient rate in near highway tends to be higher than the rate in whole area. This means that the data in big city is different from the data in whole area

(4) Algorithmic bias

This bias arises from the different outcome caused by different algorithms. For example, a search ranking by Google is different from the ranking by Bing. This means that the behaviors of users are biased by the different search engine.

(5) Interaction bias

This bias arises from the different interaction according to web presentation. When focusing on the medicine list on the pharmacy web site, for example, they are differently displayed, *e.g.*, one by one or all image. In the one by one representation, users are hard to watch the less prioritized medicines because they need to scroll the web page to find them. In the all image representation, on the other hand, users tend to watch the upper left image of medicine because we usually read the sentence from left to right and its line starts from upper to lower. Such different representations causes bias of selecting medicines.

(6) Self-selection bias

This bias arises from the different number of users who are willing to participate or not. For example, many questionnaires are returned from healthy persons, but not from the non-healthy persons. This is because healthy persons do not hesitate to tell their health information without worry about it, while non-healthy persons do not want to tell their health information honestly due to their worry about it.

(7) Second-order bias

This bias arises from the original bias. After a biased information (in the high ranking) is spread, for example, active users post other messages related to this information, and such messages are sampled in high possibility and increases its rank in search engine. This cycle amplifies the original bias.

Bias in Wellbeing AI

To clarify the bias in Wellbeing AI for easy understanding, let start to simplify the bias from the viewpoint of ML. According to Mehrabi's survey (Mehrabi et al. 2022), bias in ML arises in the cycle of (i) users, (ii) data, and (iii) algorithm as shown in Fig. 1. The connection of the seven biases on Web to the cycle from (i) to (iii) is summarized as follows. Firstly, the activity bias and self-selection bias arise in the cycle from "user" to "data" because both biases are caused by user and affect data. Secondly, the data bias and sampling bias arise in the cycle from "data" to "algorithm" because both biases are found in data and affect algorithm. Thirdly, the algorithmic bias and interaction bias arise in the cycle from "algorithm" to "user" because both biases are caused by algorithms and affect user. Finally, the second-order bias also arises in this cycle, which is the same as the bias on the web.

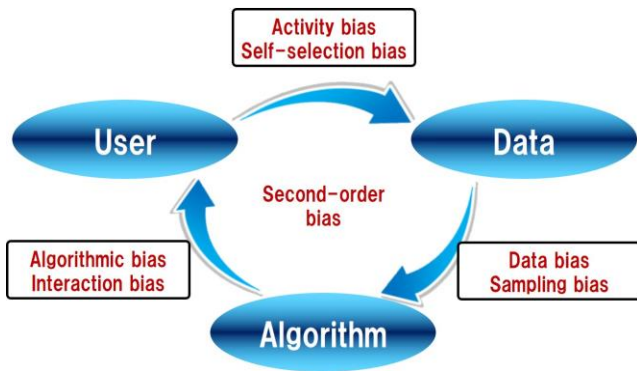


Figure 1: Bias in the cycle of ML

To consider the features of wellbeing in the cycle of arising the bias of ML that connects with the bias on Web, the following features should be taken into consideration.

● Personal information

Good information of others is not always good. For example, the knowledge of good sleep for a certain person is not always useful for other persons. This indicates that that the *personal data* is very important in wellbeing.

● Long-term evaluation

Current evaluation of health is not enough because keeping good health and better health (better life) are more important than the current health. This indicates that the *long-term evaluation* is very important in wellbeing.

From the viewpoint of the personal information and long-term evaluation, the seven biases do not arise or can be reduced as the following reasons. As shown in Figure 2, firstly, the activity bias and self-selection bias do not arise because the data comes from only one person. This indicates that the "single" user provides the "personal" data. Secondly, the data bias and sampling bias can be reduced if we can get long-term daily data. This is because such a kind of data is not heavily biased in comparison with the short-term daily data due to the large number of data. Thirdly, an influence of the algorithmic bias and interaction is very small because only one person is affected. Finally, the second-order bias can be reduced because other biases in this cycle are reduced by the above reasons. From this analysis, the algorithm in Wellbeing AI analyzes the "personal" data that comes from the "single" user and provides the result to the user. This indicates that the "personal and long-term evaluation" (precisely, the personal and long-term evaluation based on the personal data) are important for fairness AI.

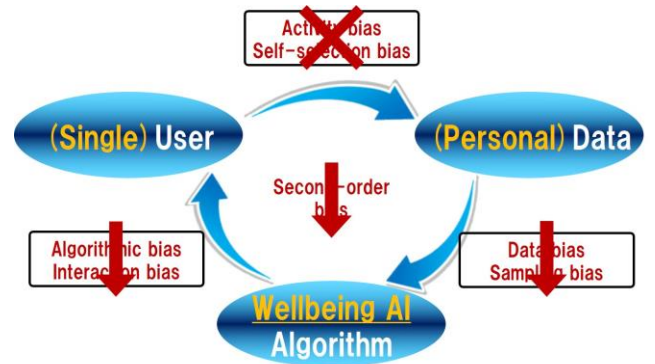


Figure 2: Bias in the cycle of Wellbeing AI

Personal and long-term evaluation

The goals of many examples of wellbeing AI are roughly classified into the following two categories.

● Keeping good health (not getting a disease)

Since many patients such as dementia, diabetes, and sleep apnea syndrome (SAS) want to worsen their health, it is important to find something wrong for early detection.

For this issue, the personal and long-term evaluation is needed for early detection.

● Better health (improving activities)

For better health, it is important to know (measure) the accumulated small progress and its change of activities of daily living (ADL) for aged persons, performance after nap for office workers, and sleep quality for all ages. For this issue, the personal and long-term evaluation is also needed.

Among them, this paper focuses on sleep of aged persons in care house because sleep is significant for aged persons. For example, aged person easily wakes up due to light sleep and may wander in midnight. Sleep can also provide some message of mind change of aged persons when their sleep quality change from good to bad. This is because persons tend to have deep sleep without anxiety but change to light sleep when they are worry about something. To tackle these issues, our previous research developed the wellbeing AI system for the first issue to extract knowledge for good sleep (Takadama et al. 2015) and for the second issue to estimate mind change of aged person from sleep (Takadama 2013). These researches took the approach of the personal and long-term evaluation (in detail, we investigated the data of the individual persons in one year). Technically, we developed the sleep quality estimation system based on vital vibration data from pressure sensor (Harada et al. 2016).

Knowledge extraction for good sleep

In our experiment, the daily activity and sleep quality are recorded every day. In one day, many activities as such meals, rehabilitation, and bathing, are scored in the integer values. For example, when eating full amount of meal, the score is 3. When no rehabilitation, then the score is 0. In addition to the dairy activity, the sleep quality (*i.e.*, the ratio of deep sleep) is estimated by our sleep stage estimation to classify whether the deep or light sleep.

Figure 3 shows the knowledge for good sleep. For person A, when the aged person had rehabilitation in AM, he became to be tired and mostly took a nap. This caused him not sleep very well at night. For this problem, our Wellbeing AI

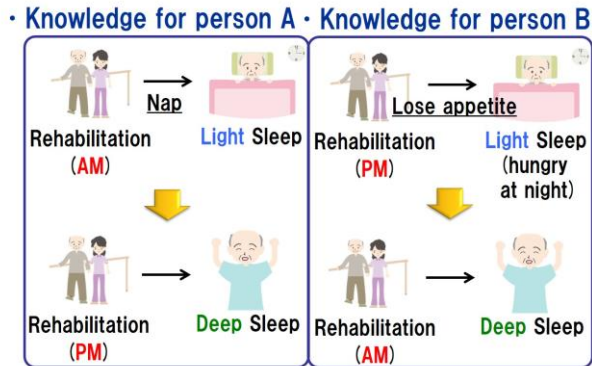


Figure 3: Knowledge for good sleep

suggested to change the time of having rehabilitation from AM to PM. As a result, he could have a deep sleep. What should be noted here is that this knowledge is not always useful for other persons. For person B, when the aged person had rehabilitation in PM, he lost appetite due to tiredness of rehabilitation and could not diner as usual. This caused him not sleep very well because of hungry at night. For this problem, our Wellbeing AI suggested to change the time of having rehabilitation from PM to AM. As a result, he could have a deep sleep. This results clearly show that the knowledge for good sleep is different among persons.

Mind change estimation of aged person

Figure 4 shows that the sleep quality of the aged diabetes person before/after the great east Japan earthquake, where the blue dots indicate the deep sleep while the red dots indicate the light sleep. The horizontal axis (f_1) indicates the achievement degree of what an aged person wants to do, while the vertical axis (f_2) indicates the achievement degree of what a care worker wants to do for an aged person. From this figure, the blue dots are located at the right side while the red dots are located at the left side before the earthquake. This is because an aged person tended to have a deep sleep when she could achieve the activities (such as eating as usual) because of not being worry about anything while she tended to have a light sleep when she is hard to achieve the activities (such as less eating as usual) because of being worry about something.

What should be noted here is that the blue and red dots were mixed after the earthquake, which had a possibility of the message of something mind changes of aged person. For this issue, our Wellbeing AI estimated that amount of breakfast should change from full to medium and the time of having rehabilitation should change from none to AM. After these changes, she could have a usual sleep. To verity these suggestions, care workers asked to her and she said that she was not willing to eat a full amount of breakfast due to news of death of many people by tsunami caused by the earthquake. Regarding the rehabilitation, she did not like it and was often absent from rehabilitation. After the earthquake, she noticed that many people killed by tsunami could not extend their life while she could extend it by having rehabilitation to tackle her diabetes. This changes her mind to be willing to exercise.

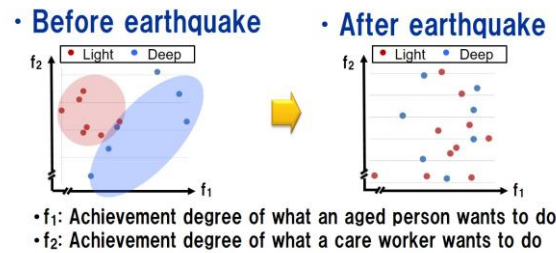


Figure 4: Light and deep sleep before/after earthquake

Conclusions

To explore the answer to the question of how we should cope with bias in Well-being AI, this paper started to focus on bias on the Internet and bias in ML and analyzed the bias in wellbeing AI after connecting biases on the Internet and bias in the ML. From this analysis, the bias in wellbeing AI can be reduced by employing the personal and long-term evaluation while many biases in ML arise. This paper discussed the fairness in Well-being AI from the viewpoint of the personal and long-term evaluation and found that our wellbeing AI based on the personal and long-term evaluation showed its potential by extracting knowledge for good sleep and estimating mind change of aged person from her sleep quality change. Future work includes an investigation of other domains.

References

- Dastin, J. 2018. "Amazon scraps secret AI recruiting tool that showed bias against women", Reuters, Oct. 11.
- Baeza-Yates, R. 2018. "Bias on the Web." *Communications of the ACM*, Vol. 61, No. 6, pp. 54-61.
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., and Galstyan, A. 2022. "A Survey on Bias and Fairness in Machine Learning." *ACM Computing Surveys*, Vol. 54, Issue 6, Article No. 115, pp. 1-35.
- Takadama, K. and Nakata, M. 2015. "Extracting Both Generalized and Specialized Knowledge by XCS using Attribute Tracking and Feedback," 2015 IEEE Congress on Evolutionary Computation (CEC2015), pp. 3034-3041.
- Takadama, K. 2013. "Towards a Care Support System that Can Guess The Way Aged Persons Feel," The AAAI 2013 Spring Symposium, Data Driven Wellness: From Self-Tracking to Behavior Change, AAAI (The Association for the Advancement of Artificial Intelligence), pp. 45-50.
- Harada, T., Uwano, F., Komine, T., Tajima, Y., Kawashima, T., Morishima, M., and Takadama, K. 2016. "Real-time Sleep Stage Estimation from Biological Data with Trigonometric Function Regression Model," The AAAI 2016 Spring Symposium, Well-Being Computing: AI Meets Health and Happiness Science, AAAI (The Association for the Advancement of Artificial Intelligence), pp. 348-353.