Introduction

The article I’m presenting is Scientists use artificial intelligence to predict biological age based on smartphone and wearables data. The term artificial intelligence here is a little bit general, and I’m going to dig deeper into the one specific technology that they use.

The recent introduction of affordable wearable sensors enables collection and storing of personal digitized activity records in cloud. This tracking is already done without interfering with the daily routines of hundreds of millions of people all over the world. I believe many of you have some form of wearable devices as well, such as Apple watch and fitbit. So many scientists have come up with the idea to utilize these records in a more informative way to monitor our health.

In March this year, which is basically two months ago, a research team from a Russian biotech company GERO and Moscow Institute of Physics and Technology has developed a computer algorithm which uses artificial intelligence and machine learning tools to predict biological age and mortality risk based on physical activities from the wearable devices that we have. The technology shows that we can combine wearable sensors and AI technologies to monitor health risks with real-time feedback to patients and wellness providers.

The team is led by Peter Fedichev, who is a russian biologist, the head of the Laboratory of Biological Systems Simulation at MIPT, and also the Science Director at GERO. Here is a quote from him: "Recent promising examples of [Artificial Intelligence] in the field of medicine include neural networks showing cardiologist-level performance in detection of the arrhythmia in ECG data, deriving biomarkers of age from clinical blood biochemistry, and predicting mortality based on electronic medical records. Inspired by these examples, we explored AI potential for Health Risks Assessment (HRA) based on human physical activity".

The product of their research is a free iOS app called GERO Lifespan that reads your health data from your phone and the wearable devices connected to your phone, such as how many steps you walk today and how long you sleep last night. It uses this information to calculate your real time life span, so that you may know, for example, whether sleeping 5 hours a day actually reduces your life span and how much will be reduced.

CNN

Brief Intro:

So today I want to focus on two questions. The first one is what algorithm do they use in exploring this AI potential to predict our lifespan and how does it actually work.

(4min)

Let’s look at the first question. The computer algorithm that they use is called convolutional neural network, which is a powerful tool to recognize patterns in machine learning. It is basically an artificial neural network that contains multiple convolution layers to detect patterns. (Each neuron receives some inputs and performs a dot product) Each convolution layer consists of many neurons that have weights and biases that are learned by the network itself during the training process. A CNN neuron is like a box that holds a value. The architecture of CNN is inspired by biological processes in that the connectivity pattern between CNN neurons resemble to the organization of animal’s neural network. We can see from this image, the input layer on the left lights up some neurons in the second column, which is the first convolutional layer and these neurons light up certain other neurons in the next column. The process continues until one final neuron in the last layer is selected. They are very much like biological neurons – an stimulus activates some groups of neurons. These firing neurons stimulate certain others to fire and output a reaction in the end.

Add an image

Deep CNN\_Age Model:

Specifically, for this research, the researchers build three age-estimating models. The first one is Deep Convolutional neural network age model, or CNN\_Age model. The other two are built to show the comparison between traditional ways of predicting age and CNN. They are regularized multivariate regression (REG\_Age model) (which is trained with a linear combination of hand-crafted features) and Principal component analysis age model (PCA\_Age model) (The paper wasn’t clear on how they build this two, but we want to focus on the comparison between these two the cnn model). We can see from the graph, the x-axis is the chronological age, or the actual age, and the y-axis is the predicted age by the three models. CNN\_Age model has the highest correlation coefficient (r) value, which indicates the age predicted by the CNN model has the strongest relationship with the chronological age and it has the highest accuracy in predicting age.

(3 min)

Let’s look at the next question. How does the CNN\_Age model actually work? The image is overview of the architecture of the CNN built by GERO. We can see they trained the network with some input data, which is physical activity data recorded by wearable devices, each contains a vector of 10080 values. Then the input enters four convolution layers and two dense layers and finally outputs one predicted age like this.

Let’s look closely into one convolution layer. (1 min)

Convolution layers are the hidden layers inside CNN. A ‘filter’ is the moving yellow box in this gif. By sliding or convolving the filter over all the input data, we compute dot products and form a ‘Feature Map’ like shown in the pink box, which is basically a matrix representation of input data. In practice, instead of we manually give values to filters, a CNN learns the values of these filters on its own during the training process.

ReLU: ReLU stands for Rectified Linear Unit. It replaces all negative values in the feature map by zero. The purpose of ReLU is to introduce non-linearity. The way I think of how ReLU works is, imagine we originally have a linear relationship like this, it will become non-linear after ReLU. since most of the real-world data would be non-linear and we want our CNN to be able to learn them as well.

The next step is max pooling. We can see how max pooling works from this picture. We first look at the 2x2 red box, and we take the largest element from the red box on the left and place it to a new matrix. Stride 2 means that we then slide our 2 by 2 window by 2 cells. Now we look at the green box and take the max again until the entire dataset is done. As a result, we get a smaller matrix than the previous one. The purpose of this operation is to reduce the size of our input representation so that the data is easier to compute and manage.

After 4 cycles of these steps, we export the result into a dense layer.

Let’s then look at a dense layer: (1 min)

A dense layer is also known as a fully connected layer. As its name suggests, every neuron in this layer is connected to every neuron of the next layer. The purpose of a dense layer is to classify the input data by the features that we detect and turn them into various classes based on the training dataset. Then ReLU function is applied again here. The next step is Dropout, it is a regularization process that prevents overfitting. If a model is overfitting, it means that it is too closely corresponded to one particular set of data. We don’t want that. We would like our model to be more general and applicable to a big range of data, that’s why we perform this kind of operation. After all these complex processes, the final output is a single linear neuron, producing the resulting value of age prediction.

(1 min)

In summary, the computer algorithm developed by GERO biotech is basically training a convolutional neural network to build an age model in these four steps. Input by Feeding physical activity records into convolutional network. Perform 4 cycles of convolution, ReLU, and max pooling. Export the result into two fully connected layers to classify them and perform other optimizations. Finally, the output layer produces an age value. These steps show how the research group generates the CNN\_Age model to accurately predict our lifespan like this, and it is also how a convolutional neural network generally works in many other researches as well. This’s all I have for today. Thank you.