Technology Reports (Special Articles) LIME Explanation Models Health Management

Special Articles on Lifestyle-enriching AI Technology

Health Risk Prediction from Health Examination Data

Service Innovation Department Taku Ito Keiichi Ochiai†1
Yusuke Fukazawa†2

From a management perspective, the idea of employee health management has become more important in recent years, to improve productivity and maintain the workforce. Promoting the health of employees is related to increasing productivity and value in an enterprise. Health risk predictions using machine learning are being implemented as a health management measure, but the learning models are black boxes, and reasons for the resulting predictions cannot be given, so it is desirable to provide convincing explanations to users. As such, NTT DOCOMO has developed models that predict health risks and can provide appropriate explanations, and a service that can make lifestyle-habit recommendations that will reduce the risk based on the predictions. The objective of the system is to raise user awareness of the relationship between bodily health and daily activity, using prediction results and convincing explanations, and to promote health by prompting changes in behavior. Users' health risks are visualized using health examination data, and we expect that promoting health will contribute to health management.

1. Introduction

In recent years, the idea of health management has increasingly been emphasized, promoting employee health from a management perspective with the objective of improving productivity and preserving the labor force. There are many initiatives to increase awareness of health, by applying

†1 Currently CrossTec Development Dept.
†2 Currently General Affairs Dept.

machine learning*1 to the results of health examinations and lifestyle habit interviews and then giving feedback on health risk predictions to the subjects. To improve awareness, employees must understand the results of prediction, so machine learning methods able to output results in a form that facilitates understanding of the reasons are needed, so that subjects can reexamine their lifestyle habits. However, most machine learning prediction models are black boxes, so it is difficult to give clear reasons that a particular health risk was predicted. There are simple models such as linear regression*2 that can be applied to prediction models to obtain explanations. These prediction and explanation models are integrated, so there is a trade-off between accuracy of prediction and simplicity of the explanation. The Explainable AI (XAI) concept has emerged to resolve this issue. One XAI method is called Local Interpretable Model-agnostic Explanations (LIME) [1], and involves prediction models and interpretation models that are trained separately.

NTT DOCOMO has used LIME to develop a health risk prediction model for employees and a service that provides lifestyle habit recommendations to reduce risk, based on the predictions. We proposed customizations to LIME to make recommendations better suited to the predictions, relative to the original version of LIME, and introduced them into the service. This article describes the proposed method as well as lifestyle habit recommendations made using the method.

2. Explainable Health Risk Prediction and Lifestyle Recommendation Technology

The proposed method for explainable health risk prediction takes the results of health examinations and lifestyle-habit interviews for a given year as input, and outputs the risk of high blood pressure and metabolic syndrome*3 N years in the future, along with the health examination items that contribute to that risk. The risk prediction method, and the explanation model used for computing the examination items that contribute to the risk are described below. We discuss explanation models including the original version of LIME and a version of LIME improved to increase the consistency of explanations, which we call consistency-emphasizing LIME.

2.1 Health Risk Prediction Algorithm

We used the eXtreme Gradient Boosting (XGBoost)*4 method for predicting health risk. We used health examination data from NTT DOCOMO employees, for whom examinations were conducted for at least four years in a row. It consists of N years of 11 examination items and 13 lifestyle-habit items as explanatory variables. We conducted training with this data to build 33 types of model to predict risk that values for each of the 11 examination items would become dangerous or not, in the three years, N+1, N+2, and N+3. Here, we used special health guidance values set by the Ministry of Health, Labour and Welfare (as of Mar. 30, 2020) as the standard to define "dangerous" values. Definitions for dangerous values are given in **Table 1**. When using Area under the ROC Curve (AuC)*5 to compare with a prediction model that simply uses the value from the previous year, predictions due to XGBoost showed improvements of from 7 to 16% in AUC.

---

*1 Machine learning: Technology that enables computers to acquire knowledge, decision criteria, behaviors, etc. from data, in ways similar to how humans acquire them, from perception and experience.

*2 Linear regression: Regression in which the relationship between the objective variable and the explanatory variable factors is linear.

*3 Metabolic syndrome: A state characterized by visceral fat obesity combined with at least two of the conditions of hypertension, hyperglycemia, and hyperlipidemia.

*4 XGBoost: A type of ensemble learning that has been attracting attention in recent years.

Table 1 Defined dangerous values for each test item

Test item	Defined dangerous values
BMI	25 or greater
Diastolic blood pressure	85 or greater
Systolic blood pressure	130 or greater
Neutral fats	149 or greater
GOT	35 or greater
GPT	35 or greater
HDL cholesterol	40 or less
LDL cholesterol	140 or greater
γ-GTP	50 or greater
Uric acid	7 or greater
Fasting blood sugar	100 or greater

2.2 Explanation Method for Health Risk Prediction Using LIME

Here we describe LIME, which provides explanations for the XGBoost prediction model. An overview of LIME is shown in **Figure 1**. With LIME, the prediction model and explanation model are trained independently. Since any algorithm can be used for the prediction model, an optimized algorithm can be used to achieve accurate results. However, as in Fig. 1, when a non-linear model*6 like XGBoost is used, for example, to predict risk of dangerous Body Mass Index (BMI) values one year later due to behaviors X and Y, it is difficult to interpret the causes directly using linear regression. Thus, we built a model to explain, for example, "what contributed to the increase in BMI

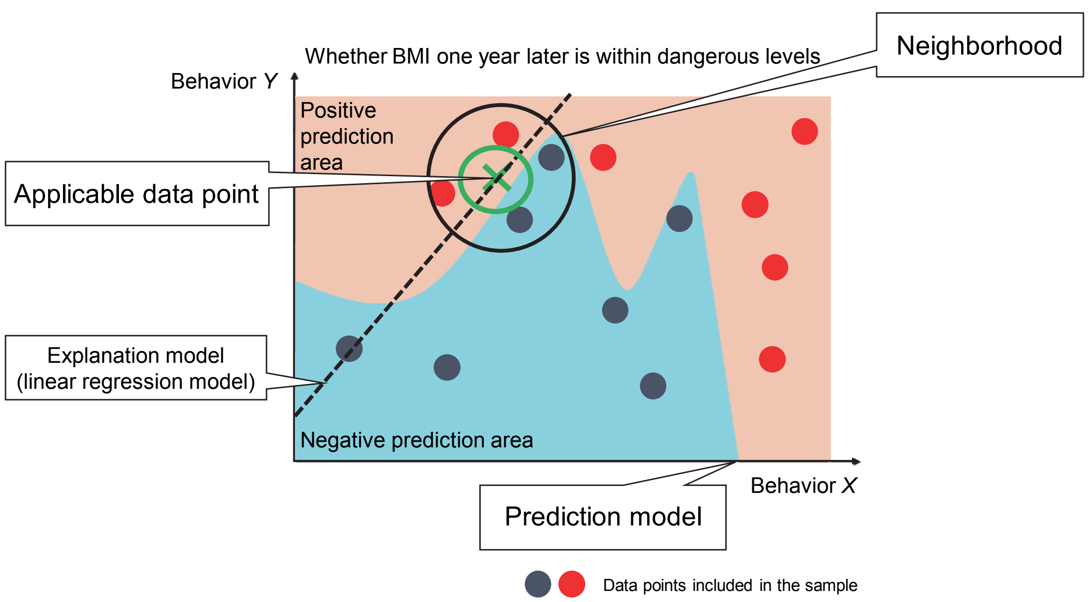

Figure 1 Overview of LIME

*5 AuC: An index for evaluating binary classifier problems. The area under the Receiver Operating Characteristic (ROC) curve will be 0.5 when predicted at random, and will be 1.0 when predictions are entirely correct.

*6 Non-linear model: A model in which the relationship between the objective variable and the explanatory variable factors is non-linear. Non-linear models generally have greater expressive power.

value?” to users for a given data point.

Explanation models extract the data point to be explained and several other nearby data points, and use them to build a linear regression model that approximates the prediction model. Since it is a linear regression model, it is possible to determine which of the explanatory variables contribute significantly, by looking at the combination of explanatory variables and partial regression coefficients*7. Since we provide an explanation using a model that approximates the prediction model near the point to be explained, the explanation provided by the model can change greatly, depending on the data points being explained. Thus, with health risk predictions, the results indicating the lifestyle habits that contribute to the health risk can differ, depending on the employee for whom the health risk is predicted, and explanations can be given for individuals.

2.3 LIME Customization Emphasizing Consistency of Explanations

An overview of consistency-emphasizing LIME is shown in **Figure 2**.

1) Issues with Conventional LIME

Since conventional LIME creates a model that approximates the data points near the points to be explained, the explanation model depends on those nearby data points. Because of this, if there are extreme data points that differ from the general tendency, there is a danger that the explanation could be distorted due to these outlying values.

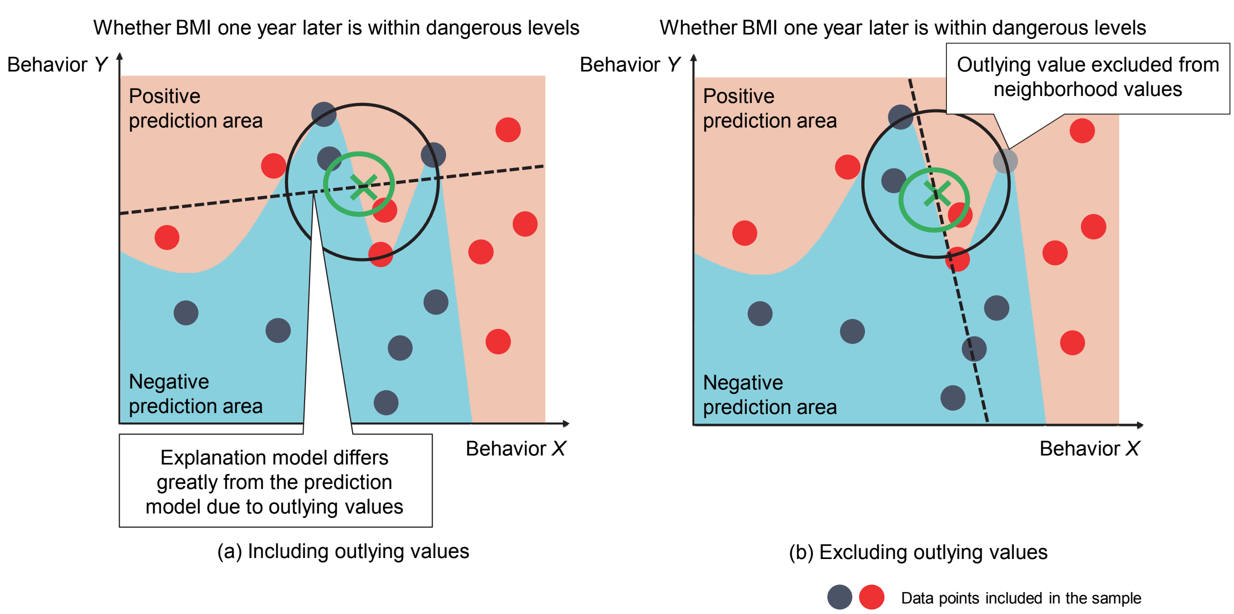

Figure 2 Overview of consistency-emphasizing LIME

*7 Partial regression coefficient: In regression analysis, a coefficient for an explanatory variable in the equation obtained.

For example, consider that there could be a sample that shows a temporary weight gain or blood pressure increase due to a disease that occurred suddenly in a given year, as well as cases where subjects became ill due to poor lifestyle habits. Since there are also cases of illness unrelated to lifestyle habits, conventional LIME could generate an explanation model that incorporates these later points since it selects nearby data at random, and this would prevent it from recommending lifestyle habit improvements correctly.

2) Consistency-emphasizing LIME

With consistency-emphasizing LIME, we have made improvements to make predictions more appropriate. Specifically, we added a mechanism that checks whether neighboring points are outlying values when searching for data points near the data point to be explained, and excludes points that are outlying values.

In performing this check, we first compute the slope of the line between the data point to be explained and a nearby data point candidate (slope A). We also compute the derivative of the prediction model at the data point to be explained (derivative B). If the characteristics of the explanatory variables match near the data point to be explained, we expect the signs of slope A and derivative B to match. For example, considering alcohol consumption and predicted BMI value one year later, if we see the characteristic, "BMI one year later increases with alcohol consumption," near the point to be explained, we expect the same tendency for other data points nearby. However, if there are outlying values, there will be data points that are significantly distant from the curve described by the prediction model. This will produce cases where the signs of slope A and derivative B do not match. As such, data points where the signs of these two values do not match are considered to be outlying values with significantly different slope than surrounding data points, and these are excluded from the sample.

A comparison of explanation models including and excluding such outlying values is shown in Fig. 2. In the figure on the left, the explanation model generated differs greatly from the prediction model due to the outlying values, but these values were excluded in the figure on the right, so the explanation model generated is close to the prediction model. This is the algorithm for consistency-emphasizing LIME.

3) Comparison of Original and Consistency-emphasizing LIME

Differences in risk prediction results for original and consistency-emphasizing LIME are shown in **Table 2**. In the table, original LIME and consistency-emphasizing LIME were applied to the same samples, the risk of metabolic disease one, two and three years later was predicted, and the lifestyle habits contributing most to that risk are compared. With original LIME, the sign of the contribution of "Eating quickly" changed between two and three years later. Normally, it would be difficult to imagine that continuing a habit of eating quickly, would yield a result that BMI decreased after two years, and then increased after three years. We can suppose that this result is due to a sample in the neighboring data points, either two years later or three years later, where a sudden change in BMI was recorded. On the other hand, with consistency-emphasizing LIME, this sort of data point is excluded, so counter-intuitive

Table 2 Change in risk prediction results between original and consistency-emphasizing LIME

<Original LIME>

Effects on BMI one-year-later		Effect on BMI two-years-later		Effect on BMI three-years-later	
Lifestyle habit	Effect score	Lifestyle habit	Effect score	Lifestyle habit	Effect score
Alcohol consumed	−0.14	Physical activity	−0.17	Increased body weight	0.08
Alcohol habit	0.09	Eating quickly	−0.08	Eating quickly	0.03
Physical activity	−0.06				

Predictions are not consistent

<Consistency-emphasizing LIME>

Effects on BMI one-year-later		Effect on BMI two-years-later		Effect on BMI three-years-later	
Lifestyle habit	Effect score	Lifestyle habit	Effect score	Lifestyle habit	Effect score
Eating dinner before going to bed	0.05	Eating dinner before going to bed	0.04	Sleep	0.03
Alcohol habit	−0.03	Physical activity	−0.04	Alcohol consumed	0.03
		Sleep	0.03		

predictions like that described above do not occur. This sort of validity is important for presenting risk prediction results that are convincing.

3. Lifestyle-habit Recommendation Technology Service Development

The technology described above was developed into a service on NTT DOCOMO d-Healthcare, and released in April last year as a service that enables NTT Group employees to check their own health risks and also lifestyle habits that contribute to those risks. A screen shot of the service is shown in **Figure 3**.

To use the service, members of the NTT Health Insurance Union first give permission for results of their own health examinations registered in "NTT Health Portal Navi" to be transferred to d-Healthcare. A library[*8] that performs the computations described above is then applied to the data, to output health risks and the lifestyle habits contributing to those risks. The model used in the library was trained using over 10,000 health-check data points collected by NTT DOCOMO over several years.

4. Conclusion

As the emphasis on health management is increasing, in this article, we have described a method using machine learning to predict health risks, the difficulty in providing explanations for such predictions, and an approach that solves this issue.

An actual service for company employees was

*8 **Library**: A collection of versatile programs in a reusable form.

Figure 3　d-Healthcare service screenshot

developed, and we have begun offering it to companies outside of NTT, as a service to support health management in corporations. In the future, we plan to continue improving the technology, contributing to improving employee health, and promoting NTT DOCOMO's health management outside of the company.

REFERENCES

[1] M. T. Ribeiro, S. Singh and C. Guestrin: ""Why Should I Trust You?": Explaining the Predictions of Any Classification," KDD'16, 2016.

[2] Association for Preventative Medicine of Japan: "Examining health examination results: Specific health examination (Metabolic syndrome)," (In Japanese). https://www.jpm1960.org/exam/exam01/exam15.html

* Data used here is from Health Data Bank. Explanation of how data would be used was given and permission from subjects was obtained before data was collected.