

Extracting Lifestyle Rules for Reduction of Body Fat Mass Using Inductive Logic Programming

Sho Ushikubo, Katsutoshi Kanamori, and Hayato Ohwada

Abstract—This study was performed to extract rules for reducing body fat mass for preventing lifestyle-related diseases. Lifestyle-related diseases have been increasing in Japan, even among younger people. Body fat mass is related to lifestyle-related diseases. Hence, finding rules for reducing body fat mass is very meaningful. We obtained lifestyle time-series data from five male subjects who are in their 20s and not obese. The data includes the amount of body fat mass in each subject and a variety of features. We used Inductive Logic Programming (ILP) to apply this data because ILP can more flexibly learn rules than other machine-learning methods. As a result of applying the data to ILP, our ILP system has successfully extracted rules to decrease body fat mass based on limited data. Learned rules indicate that a combination of sufficient sleep and low intake of carbohydrates; a combination of duration of protein and low intake of fat; or a combination of sufficient protein intake, low fat intake, and sufficient vitamin D intake is the most effective lifestyle in decreasing body fat mass. Earlier studies have amply demonstrated that nutrient intake alone is effective in decreasing body fat mass using statistical analysis. However, this study revealed a new finding that a combination of sufficient sleep and low intake of carbohydrates is also effective, and we can infer that a combination with some nutrients is more effective in decreasing body fat mass than nutrients alone.

Index Terms—Inductive logic programming, knowledge discovery and data mining, machine learning, lifestyle-related diseases.

I. INTRODUCTION

This chapter describes the objective of this study and related works. The change in health and social lifestyles in recent years has led to increasing numbers of patients suffering from lifestyle-related diseases, even among younger people, posing important social problems in Japan [1]. These lifestyle-related diseases are chronic diseases including cancer, heart diseases, and diabetes. Nutrition and dietary habits, physical activities, and sleep are important factors that influence lifestyle-related diseases. Moreover, body fat mass is related to lifestyle-related diseases. Hence, we should encourage having an appropriate body fat mass when a person is young to prevent lifestyle-related diseases. However, all factors that decrease body fat mass have not yet been clarified.

Excess energy intake for a long period leads to an increase in body weight. Additionally, earlier studies have amply demonstrated that certain nutrients have the effect of decreasing body fat mass and weight. Robert *et al.* [2] and

Zemel *et al.* [3] suggest that dietary calcium may play a role in the regulation of body fat and weight using statistical analysis. Similarly, Halton *et al.* [4] show that high protein diets have a potential for weight loss. Moreover, Wei *et al.* [5] suggest that consuming a combination of calcium and vitamin D may be associated with lower body weight and better metabolic health. These clinical studies consider the obesity state of the subjects and specific nutrients only.

We consider a variety of factors such as other nutrients, exercise, and sleep. If a variety of factors are considered, there is a limit to the degree to which relationships between decreasing body fat mass or weight and those factors can be correlated using statistical analysis. This study therefore used Inductive Logic Programming (ILP) [6]. ILP can easily and logically express the relationships among complex features. Other classification methods featuring machine learning such as SVM [7] or Random Forest [8] can handle complex features with high accuracy, but they are best in cases where we do not understand what the features do or what they mean. In contrast, ILP can derive rules in forms that we can easily understand. Decision Trees [9] can also derive rules in forms that we can easily understand, but ILP has the advantages that it can learn rules flexibly using background knowledge represented by predicate logic, and that it can discover rules from a multi-relational database consisting of multiple tables.

Therefore, the objective of this study was to extract lifestyle rules for reducing body fat mass from time-series lifestyle data that include many features using ILP. Data were obtained for five male subjects who were not obese. The data includes the body fat mass of each subject. We process the data in a form that ILP can understand. We then apply it to ILP and extract rules to decrease the body fat mass. We believe that when we live based on these learned rules, we can maintain an appropriate body fat mass.

II. DATASET

This chapter describes the dataset. In this study, time-series lifestyle data were obtained every day from September 17, 2015, to October 31, 2015. In this study there were five male subjects who were in their 20s and not obese. We chose only men as subjects because there are more obese men than women in Japan. Note that we assessed whether the subjects were obese or not based on BMI. Table I lists the number of daily lifestyle data points for each subject.

TABLE I: NUMBER OF DAILY LIFESTYLE DATA POINTS FOR EACH SUBJECT

Subject ID	Number of data points
1	42
2	44
3	45
4	31
5	45

Manuscript received December 5, 2015; revised February 15, 2016.

Sho Ushikubo, Katsutoshi Kanamori, and Hayato Ohwada are with Tokyo University of Science, Noda city, Chiba prefecture, Japan (e-mail: 7415605@ed.tus.ac.jp, katsu@rs.tus.ac.jp, ohwada@rs.tus.ac.jp).

The daily body fat mass, sleeping hours, sleep quality, bedtime, stress, calorie consumption, and daily food intake for each subject are included in the obtained data (Fig. 1). Basal metabolism is also included in calorie consumption. Sleep quality and stress are qualitative variables, with each subject choosing from -1 or 0, or 1. In sleep quality, -1 represents “Unsatisfied,” 0 represents “Cannot say,” and 1 represents “Satisfied.” Similarly, in stress, -1 represents “Feel little stress,” 0 represents “Cannot say,” and 1 represents “Feel much stress.” The other features are continuous variables. The daily calorie consumption and sleeping hours are provided by wearable devices.

Subject id	Date	Body fat mass	Sleeping hours	Sleep quality	Bedtime	Stress	Calorie consumption	Food name	Amount	...
1	9/17	11.4	7	0	29	0	2143	Egg	30	...
2	9/17	6.95	7	1	25	-1	2344	Beef	100	...
⋮										
5	10/31	11.3	4	1	27	0	1907	Chicken	200	...

Daily sleep data
Daily food intake data

Fig. 1. Obtained time-series of raw lifestyle data.

III. METHOD

This chapter describes our method and the way we process time-series data.

A. ILP

This study used ILP. ILP combines machine learning and logic programming. Given a formal encoding of the background knowledge, positive examples, and negative examples, an ILP system will derive hypotheses that explain the positive examples and negative examples. In this approach, logic is used as a language to induce hypotheses from the examples and background knowledge. Thus, the result of learning is represented as a logical formula in predicate logic.

Compared with other machine-learning approaches, ILP has several advantages. First, ILP is able to discover knowledge from a multi-relational database consisting of multiple tables. We cannot do this using a decision tree. Second, using logic programming enables us to encode more general forms of background knowledge such as recursions, functions, or quantifiers. Finally, the learned rules are comprehensible by humans.

In this study, we used GKS [10] as our ILP system software.

B. Data Processing

An ILP system cannot deal with a continuous variable. Hence, we need to convert the raw data into qualitative form. Using a Cookpad data set provided by Cookpad and the National Institute of Informatics, the Standard Tables of Food Composition in JAPAN -2010- [11], and the National Nutrient Database for Standard Reference Release 28 [12], we converted the daily food intake data recorded by the subjects into the daily nutrient intake data seen in Fig. 2.

Table II shows 31 features of nutrition intake incorporated in Fig. 2.

Moreover, we transformed the arranged time-series lifestyle data into qualitative form so that ILP could understand it, using the following steps.

Step 1: ΔB is defined as the difference between the body fat mass on a given day and that for the day before.

Step 2: We decide that the class is 1 (Decrease) if $\Delta B \leq 0$ for that day and ID, and that the class is 0 (Increase) if $\Delta B > 0$.

Step 3: We translate the continuous features into corresponding qualitative values using categories (e.g., average, low, middle, or high) based on the quantiles in the standard normal distribution.

Subject id	Date	Body fat mass	Sleeping hours	Calorie consumption	Calories	Salt equivalent
1	9/17	11.4	7	2143	1313	7.92
2	9/17	6.95	7	2344	1194	10.2
⋮						
5	10/31	11.3	4	1907	1706	8.14

31 features of nutrition intake

Fig. 2. Arranged time-series lifestyle data.

TABLE II: 31 FEATURES OF NUTRITION INTAKE

Feature	Description
Calories	Calorie intake [kcal]
Protein	Protein intake [g]
Fat	Fat intake [g]
Carbohydrate	Carbohydrate intake [g]
Vitamin A	Vitamin A intake [μg]
Vitamin D	Vitamin D intake [μg]
Vitamin E	Vitamin E intake [mg]
Vitamin K	Vitamin K intake [μg]
Vitamin B1	Vitamin B1 intake [mg]
Vitamin B2	Vitamin B2 intake [mg]
Niacin	Niacin intake [mg]
Vitamin B6	Vitamin B6 intake [mg]
Vitamin B12	Vitamin B12 intake [μg]
Folate	Folate intake [μg]
Pantothenic acid	Pantothenic acid intake [μg]
Biotin	Biotin intake [μg]
Vitamin C	Vitamin C intake [mg]
Potassium	Potassium intake [mg]
Calcium	Calcium intake [mg]
Magnesium	Magnesium intake [mg]
Phosphorus	Phosphorus intake [mg]
Iron	Iron intake [mg]
Zinc	Zinc intake [mg]
Copper	Copper intake [mg]
Manganese	Manganese intake [mg]
Iodine	Iodine intake [μg]
Selenium	Selenium intake [μg]
Chromium	Chromium intake [μg]
Molybdenum	Molybdenum intake [μg]
Dietary fiber	Dietary fiber intake [g]
Salt equivalent	Salt equivalent intake [g]

The final lifestyle time-series data are presented in Fig. 3.

Subject id	Date	ΔB	Sleeping hours	Calorie consumption	Calories	Salt equivalent
1	9/18	1	ave	low	low	ave
2	9/18	0	ave	ave	low	low
⋮						
5	10/31	1	low	mid	mid	low

Step 1 and Step 2

Step 3

Fig. 3. Lifestyle time-series data that ILP can understand.

C. Background Knowledge and Examples

Table III presents a set of predicate types and their mode

declarations representing the background knowledge. Some parts of the predicates used to represent the features are shown below. All predicates are described by a subject ID, a date, and a parameter value. The + and # symbols indicate input variables and constants. Other features such as calorie consumption and calories are also expressed in this way.

Sleeping _hours(+Id _and _date,#Day _before,#Val)
Sleep _quality(+Id _and _date,#Day _before,#Val)

For example, in *Sleeping _hours*, we can express the lifestyle time-series data in Fig. 3 as follows:

Sleeping _hours(+1 _and _9/30,#0,#ave)
Sleeping _hours(+1 _and _9/30,#2,#ave)

The upper expression represents the qualitative value of *Sleeping hours* for subject ID 1 on September 30. The lower expression represents the qualitative value of *Sleeping hours* for subject ID 1 on September 28. Specifically, *#Day _before* in the lower expression represents two days before September 30, i.e., September 28, and is set to express a time-series. The range of *#Day _before* is from 1 to 3.

By selecting several predicates as background knowledge in Table III, we can extract rules as a learning result. We select several predicates because if we set all predicates as background knowledge, it takes a long time and an out-of-memory error eventually occurs.

TABLE III: PREDICATES SELECTED FOR BACKGROUND KNOWLEDGE

Set of predicates	Description
Sleeping hours, Sleep quality, Bedtime, Stress, Calorie consumption, Protein, Fat, and Carbohydrate	Consider sleep, stress, exercise and three major nutrients
Vitamin D, Vitamin E, Vitamin B1, Vitamin B2, Niacin, Vitamin B6, Vitamin B12, Pantothenic acid, and Biotin	Consider vitamins
Potassium, Calcium, Magnesium, Phosphorus, Iron, Zinc, Manganese, Iodine, Selenium, Chromium, and Molybdenum	Consider minerals
Sleeping hours, Sleep quality, Bedtime, Calorie consumption, Protein, Fat, Carbohydrate, Calcium, and Vitamin D	Add calcium and vitamin D based on related work [5]

In this study, there were 109 positive examples (class 1) and 85 negative examples (class 0)

D. Model Evaluation

We use accuracy, precision, recall and F-measure. Accuracy, precision, recall, and F-measure are calculated as (1), (2), (3), and (4).

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{F-measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

TP, FP, FN, and TN in these formulas represent True

Positive, False Positive, False Negative, and True Negative, and these are related as seen in Table IV.

TABLE IV: DEFINITIONS OF TP, FP, FN, AND TN

		observed	
		1	0
Predicted	1	TP	FP
	0	FN	TN

IV. RESULTS AND DISCUSSION

The ILP results are listed in Table V, in which the set for sleep, exercise and three major nutrients (predicate set A) is the most effective, and the set for vitamin D and calcium (predicate set B) is the second-most effective based on the F-measure.

TABLE V: ILP RESULTS FOR EACH SET OF PREDICATES

TP	FP	FN	TN	Accuracy	Precision	Recall	F-measure
89	59	20	26	0.593	0.601	0.817	0.693
87	67	22	18	0.541	0.565	0.798	0.662
87	61	22	24	0.572	0.588	0.798	0.677
90	66	19	19	0.562	0.577	0.826	0.680

In addition, we present rules for predicate set A based on the minimum description length principle [13]. {T, F} denotes the number of positive examples (T) and the number of negative example (F) that the rule covers.

Rule 1: {18, 1} class(A) :-Sleeping_hours(A, 3, low),
 Carbohydrate(A, 1, mid)

Rule 2: {18, 0} class(A) :-Protein(A, 3, low),
 Protein(A, 1, low), Fat(A, 2, mid)

Rule 1 indicates that sufficient sleep for three days before and low carbohydrate intake on the day before are related to decreasing body fat. Short sleep duration is related to increasing the total fat intake [14], and a short-term low carbohydrate diet is related to decreasing body fat [15]. Hence, this rule can be naturally understood. However, there was a new finding revealing that the combination of sufficient sleep and low carbohydrate intake is effective in decreasing body fat mass. Moreover, Rule 2 indicates that protein duration and low fat intake are related to decreasing body fat. We can also naturally understand this rule because it is known that daily protein intake helps us lose weight, and that standard-protein and low-fat diets are effective in losing body weight [16].

Similarly, we present rules for predicate set B.

Rule 3: {18, 2} class(A) :-Protein(A, 3, low), Fat(A, 2, mid),
 Vitamin D(A, 0, low)

Rule 3 indicates that sufficient protein intake three days before, low fat intake two days before, and sufficient vitamin D intake are related to decreasing body fat. Previous studies have shown that vitamin D deficiency is a cause of obesity [17], [18], and that standard-protein and low-fat diets are effective, as described above [16]. However, new findings reveal that the combination of sufficient protein intake, low fat intake, and sufficient vitamin D intake is effective in decreasing body fat mass. In earlier research [5], we extracted the rule that vitamin D and calcium are related to decreasing

body fat mass, but the number of positive and negative examples of these rules is 13 and 1, respectively. Hence, Rule 3 covers more examples than the rule about vitamin D and calcium.

Accordingly, we extracted rules with high F1 scores from few data, and it is very probable that lifestyle time-series data can be applied to ILP. Earlier studies have used statistical analysis to amply demonstrate that a certain nutrient intake is effective in decreasing body fat mass. However, based on this study, we can infer that a combination of the intake of certain nutrients and lifestyle factors such as sleep and exercise are more effective in decreasing body fat mass than the nutrients alone. In addition, we succeeded in expressing the duration of nutrients using predicate logic rather than a statistical analysis. If we live based on the learned rules, we believe we will be able to maintain an appropriate body fat mass.

V. CONCLUSION AND FUTURE WORK

This paper applied ILP to extract rules for reducing body fat mass using lifestyle time-series data. We described how to acquire the raw data, arrange the raw data, and transform the data into qualitative data that ILP can understand. We then constructed background knowledge and examples, and learned rules to decrease body fat mass. The learned rules indicate that a combination of sufficient sleep and low carbohydrate intake, or a combination of protein duration and low fat intake or a combination of sufficient protein intake, low fat intake, and sufficient vitamin D intake is effective in decreasing body fat mass. In addition, we succeeded in expressing the duration of nutrients using ILP. We hope that ILP will be useful for preventing lifestyle-related diseases and assisting with weight or fat-mass control.

Future work includes investigating the applicability of the learned rules to other subjects. We selected male subjects in their 20s who were not obese, but we have to consider females, people of all ages, and obese people. Furthermore, we should try other combinations of predicates, such as using several vitamins and minerals.

ACKNOWLEDGMENT

In this paper, we used recipe data provided by Cookpad and the National Institute of Informatics.

REFERENCES

- [1] Ministry of Health, Labour and Welfare. Patient Survey. (Oct. 2011). [Online]. Available: <http://www.mhlw.go.jp/toukei/saikin/hw/kanja/11/index.html>
- [2] R. P. Heaney, K. M. Davies *et al.*, "Calcium and weight: Clinical studies," *Journal of the American College of Nutrition*, vol. 21, no. 2, pp. 152-155, Apr. 2002.
- [3] M. B. Zemel, H. Shi, B. Greer, D. Dirienzo, and P. C. Zemel, "Regulation of adiposity by dietary calcium," *FASEB Journal Trans*, vol. 14, no. 9, pp. 1132-1138, Jun. 2000.
- [4] T. L. Halton and F. B. Hu, "The effects of high protein diets on thermogenesis, satiety and weight loss: A critical review," *J Am Coll Nutr*, vol. 23, no. 5, pp. 373-385, Oct. 2004.
- [5] W. Zhu, D. L. Cai *et al.*, "Calcium plus vitamin D3 supplementation facilitated Fat loss in overweight and obese college students with very-low calcium consumption: A randomized controlled trial," *Nutrition Journal*, vol. 12, no. 8, pp. 12-43, Apr. 2013.
- [6] S. Muggleton, "Inductive logic programming," *New Generation Computing*, vol. 8, no. 4, pp. 295-318, Feb. 1991.
- [7] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. the 5th Annual ACM Workshop on Computational Learning Theory*, July 1992, pp. 144-152.
- [8] Leo. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, Oct. 2001.
- [9] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, no. 1, pp. 81-106, 1986.
- [10] H. Ohwada, H. Nishiyama, and F. Mizoguchi, "Concurrent execution of optimal hypothesis search for inverse entailment," *Lecture Notes in Artificial Intelligence*, vol. 1866, no. 4, pp. 165-173, Aug. 2000.
- [11] The Ministry of Education, Culture, Sports, Science and Technology-Japan. (May 2014). *Standard Tables of Food Composition in JAPAN -2010-*. [Online]. Available: http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu3/houkoku/1298713.htm
- [12] United States Department of Agriculture Agricultural Research Service. (Dec. 2011). *National Nutrient Database for Standard Reference Release 28*. [Online]. Available: <http://ndb.nal.usda.gov/ndb/search>
- [13] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465-471, Sep. 1978.
- [14] H. S. Dashti *et al.*, "Short sleep duration and dietary intake: Epidemiologic evidence, mechanisms, and health implications," *Advances in Nutrition*, vol. 6, pp. 648-659, Nov. 2015.
- [15] A. J. Nordmann, A. Nordmann *et al.*, "Effects of low-carbohydrate vs low-fat diets on weight loss and cardiovascular risk factors: A meta-analysis of randomized controlled trials," *Arch Intern Med*, vol. 166, no. 3, pp. 285-293, Feb. 2006.
- [16] D. P. Belobrajdic, J. Frystyk *et al.*, "Moderate energy restriction-induced weight loss affects circulating IGF levels independent of dietary composition," *European Journal of Endocrinology*, vol. 162, no. 6, pp. 1075-1082, Jun. 2010.
- [17] Y. J. Foss, "Vitamin D deficiency is the cause of common obesity," *Medical Hypothesis*, vol. 72, no. 3, pp. 314-321, March 2009.
- [18] J. Wortsman, L. Y. Matsuoka, T. C. Chen, Z. Lu, and M. F. Holick, "Decreased bioavailability of vitamin D in obesity," *Am J Clin Nutr*, vol. 72, no. 3, pp. 690-693, Sep. 2003.



Sho Ushikubo graduated from the Department of Industrial Administration, Faculty of Science and Technology, Tokyo University of Science, Noda City, Japan, in 2015.

He is a student at Tokyo University of Science Graduate School, Division of Science and Engineering Industrial Administration Master's Course since 2015, Noda City, Japan. His research interests are in the field of machine learning.



Katsutoshi Kanamori earned a doctorate in information science at the Tokyo University of Science, Noda City, Japan, in 2009.

He is a research associate at Tokyo University of Science, Department of Industrial Administration, Faculty of Science and Technology. His research interests are in the fields of artificial intelligence.



Hayato Ohwada graduated from the Department of Industrial Administration, Faculty of Science and Technology, Tokyo University of Science, Noda City, Japan, 1983. Then he graduated from Tokyo University of Science Graduate School, Division of Science and Engineering Industrial Administration Doctoral course Completed program with degree, Noda City, Japan, 1988.

He was a research associate (Tokyo University of Science) from 1988 to 1998, lecturer (Tokyo University of Science) from 1999 to 2000, and associate professor (Tokyo University of Science) from 2001 to 2004. Then he is a professor at Tokyo University of Science Faculty of Science and Engineering Department of Industrial Administration from 2005. His research interests are in the fields of inductive logic programming and bioinformatics.