P309

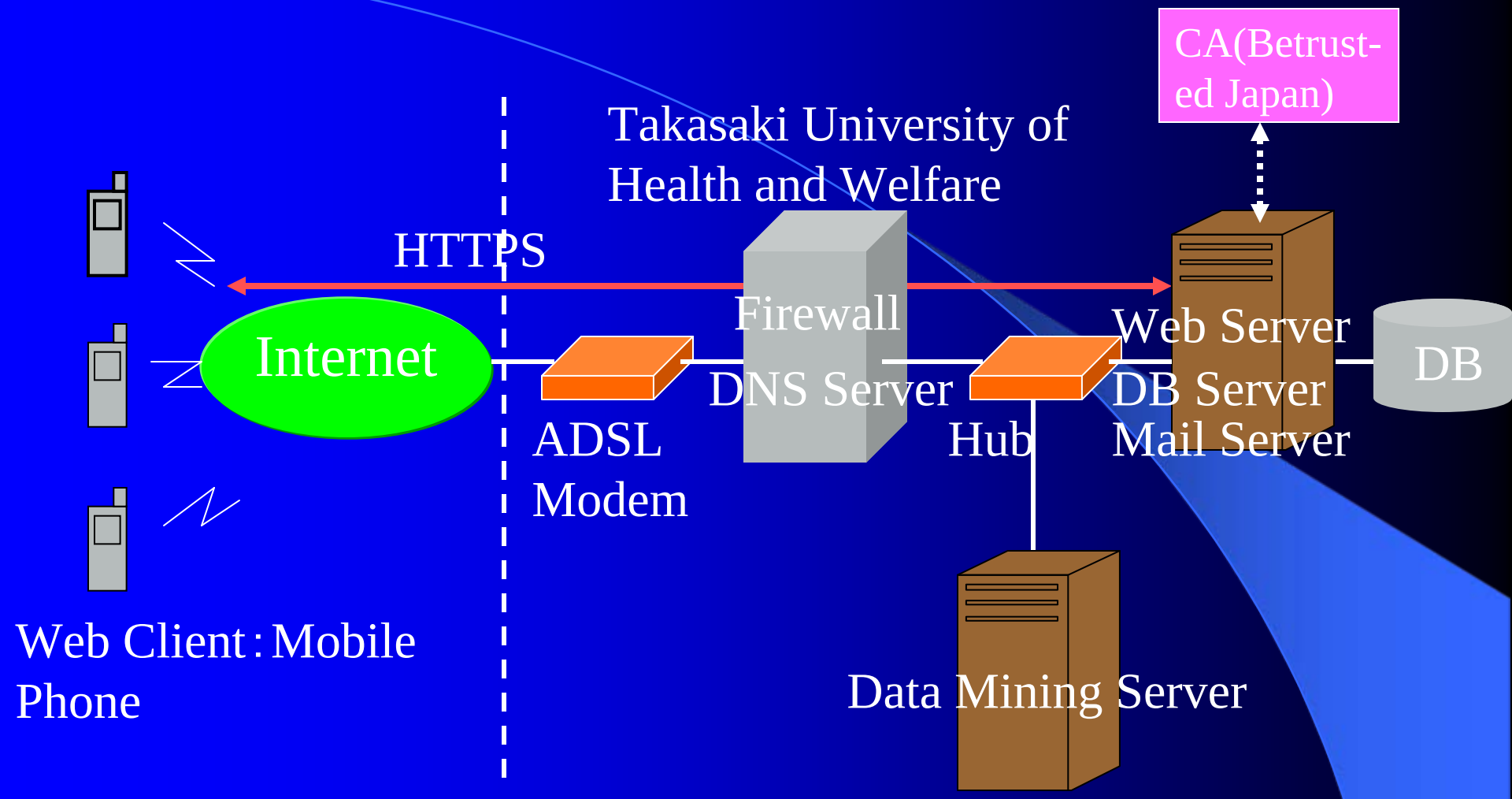
Time-Series Data Analyses for Healthcare-Data-Mining Based on a Personal Dynamic Healthcare System

Hiroshi Takeuchi [1)], Yukari Ikeda[2)]

Naoki Kodama [1)]

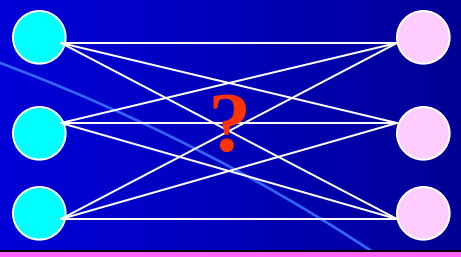
[1)]Department of Healthcare Informatics, Takasaki University of Health and Welfare, Takasaki, Gunma, Japan

[2)]Ota General Hospital, Ota, Gunma, Japan



Configuration of personal dynamic healthcare system

lifestyle data
(input fields)



health data
(output fields)

Extracting personal rules relating lifestyle and health data

ingestion

energy expenditure

sleep

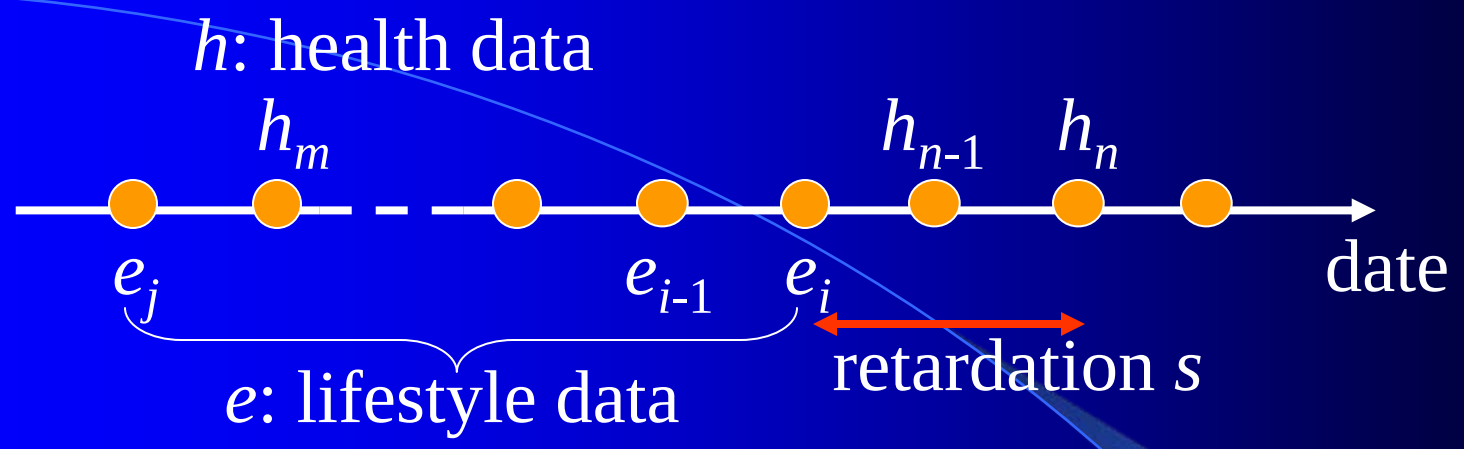
alcohol intake

smoking

daily vital signs

(blood pressure, body-fat index, body-temperature etc.)

Concept of healthcare-data-mining



$$\Delta h_{nm} = h_n - h_m$$ (variation of health data)

Correlation check

$$e_{ij} = e_i + e_{i-1} + \ldots + e_j$$ (summation of lifestyle data)

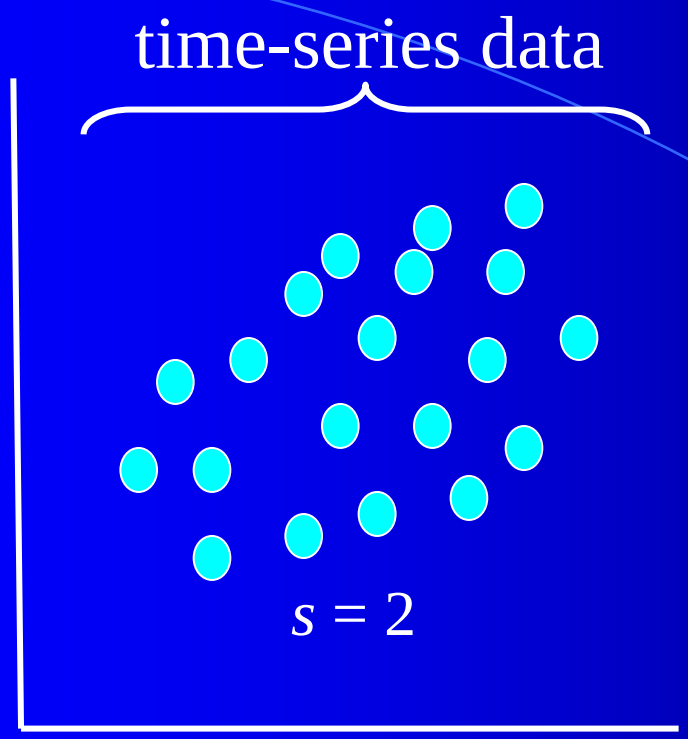
$$(i <= n - 1 : s = n - i)$$

Calculate correlation between Δh_{nm} and e_{ij} in time-series data while changing *n-m*, *i-j*, and *s* as parameters

Algorithm for defining the input field (1)

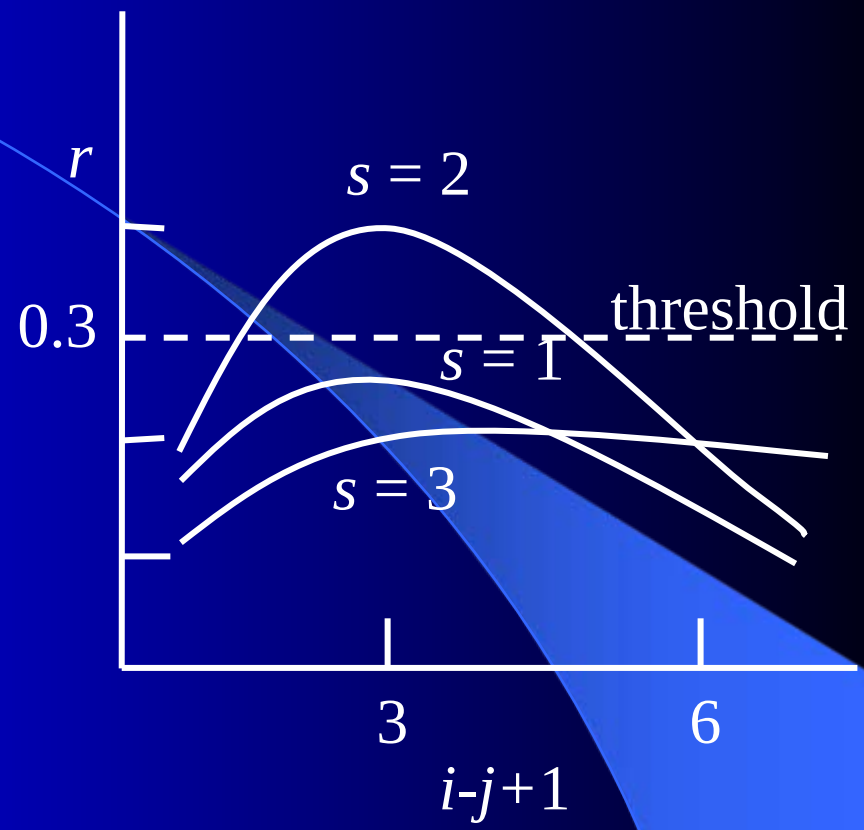
Δh_{nm}

(n-m = 3)



$$e_i + e_{i-1} + e_{i-2}$$

(i = n-2)



r : correlation coefficient

input field is $e_i + e_{i-1} + e_{i-2}$ (i=n-2) in this case

Algorithm for defining the input field (2)

Data set

Date	Energy expenditure 5 (kcal)	Ingestion 5 (kcal)	Alcohol intake	Body-fat Index
	Y_i			X
06/06	1000	12000	much	higher
06/07	1200	11000	little	lower
06/08	800	13500	moderate	moderate
....	....	....	....	....

Rule

If Y_1 is y_1 and Y_2 is y_2, then $X=x$ with probability p.

Y_i : lifestyle data item, y_i : condition X : health data item, x : value

Rules having higher-J-measure survive in the rule table

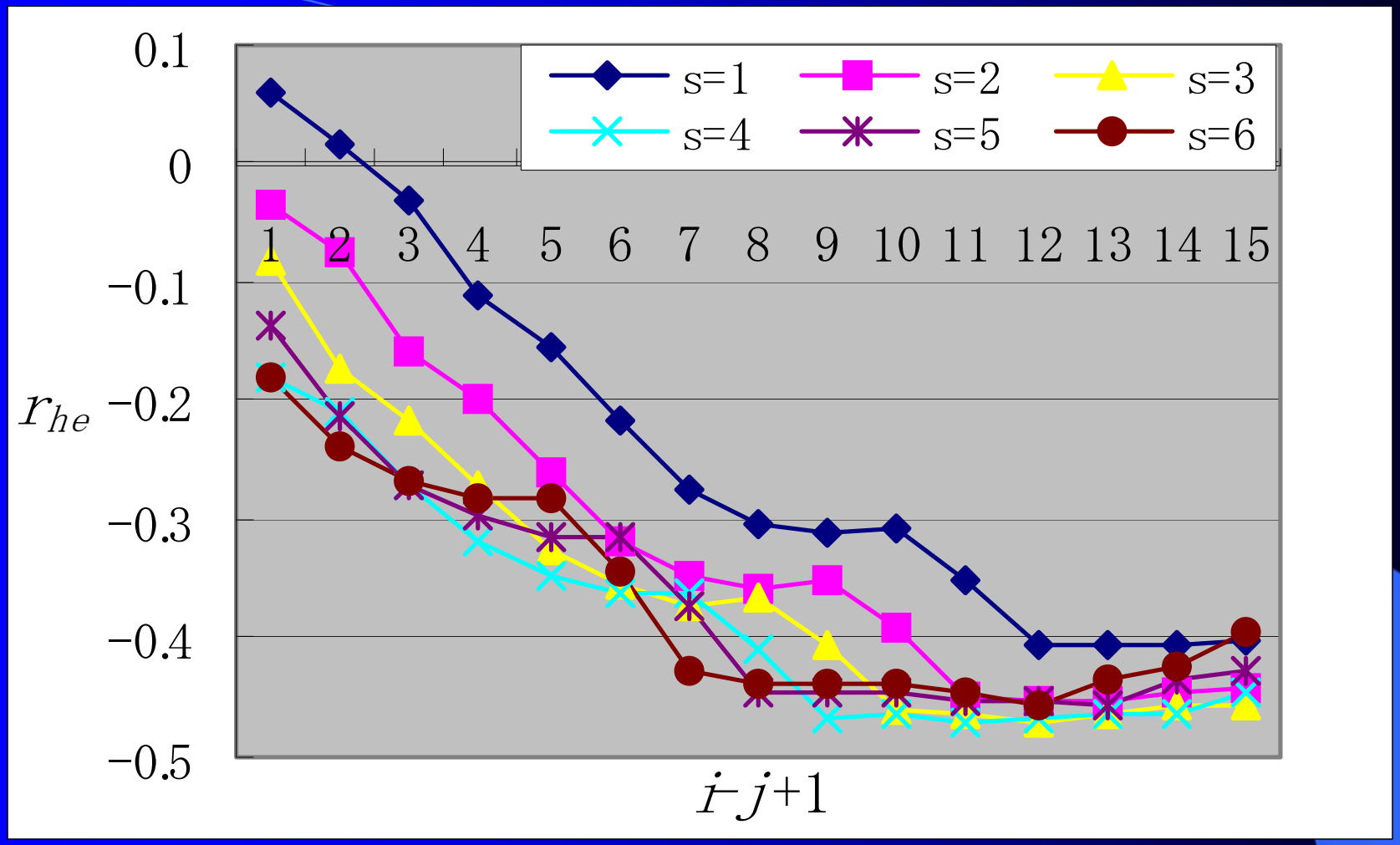
J-measure

$$J(x|y) = p(y)\left(p(x|y)\log\frac{p(x|y)}{p(x)} + (1-p(x|y))\log\frac{(1-p(x|y))}{(1-p(x))} \right)$$

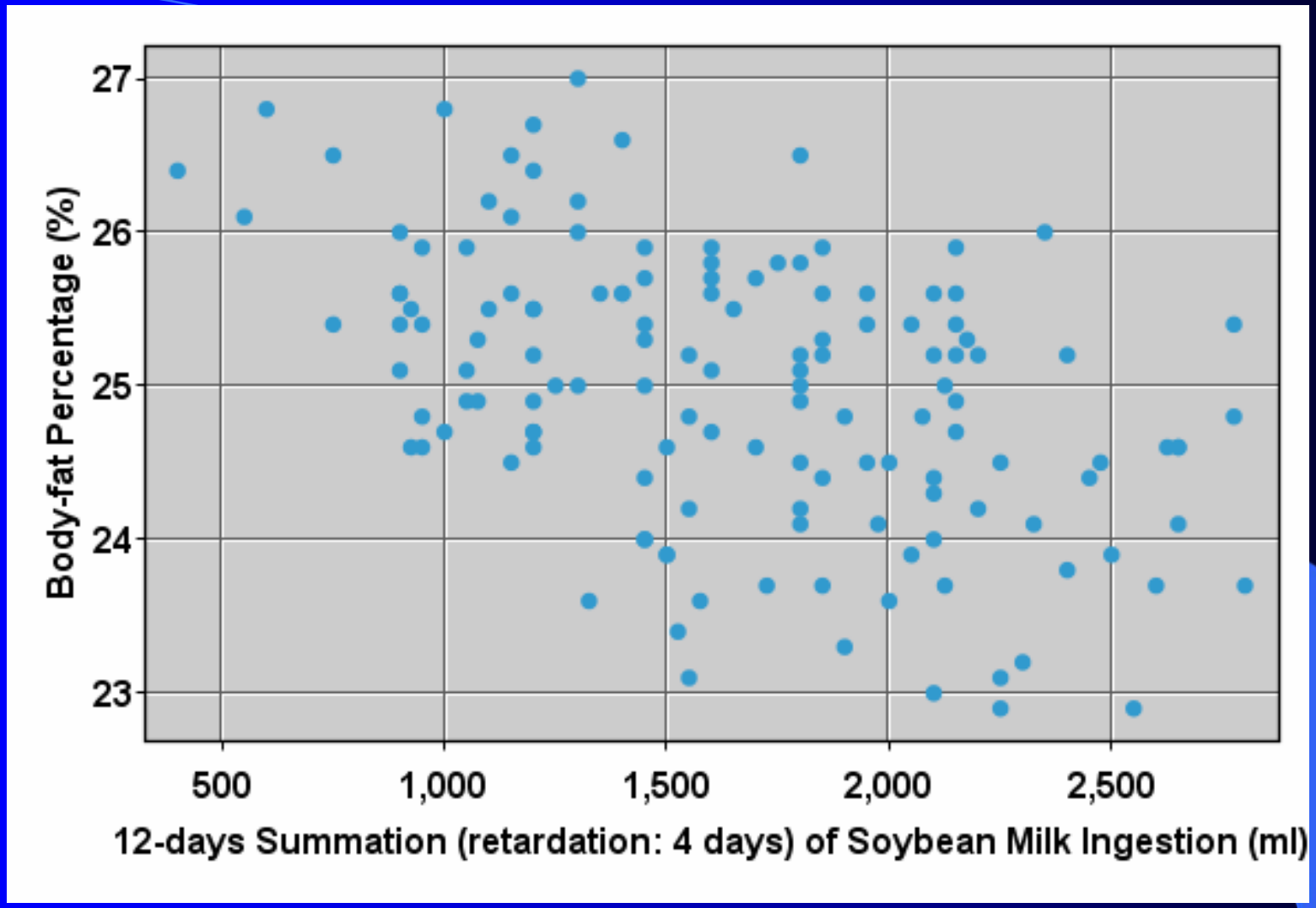
$\left.\begin{matrix} p(y) \\ p(x) \end{matrix}\right\}$ probability of the rule's $\left\{\begin{matrix} \text{antecedent} \\ \text{consequent} \end{matrix}\right\}$ matching an example from the data set

$p(x|y)$:conditional probability of the rule's consequent conditioned on the antecedent

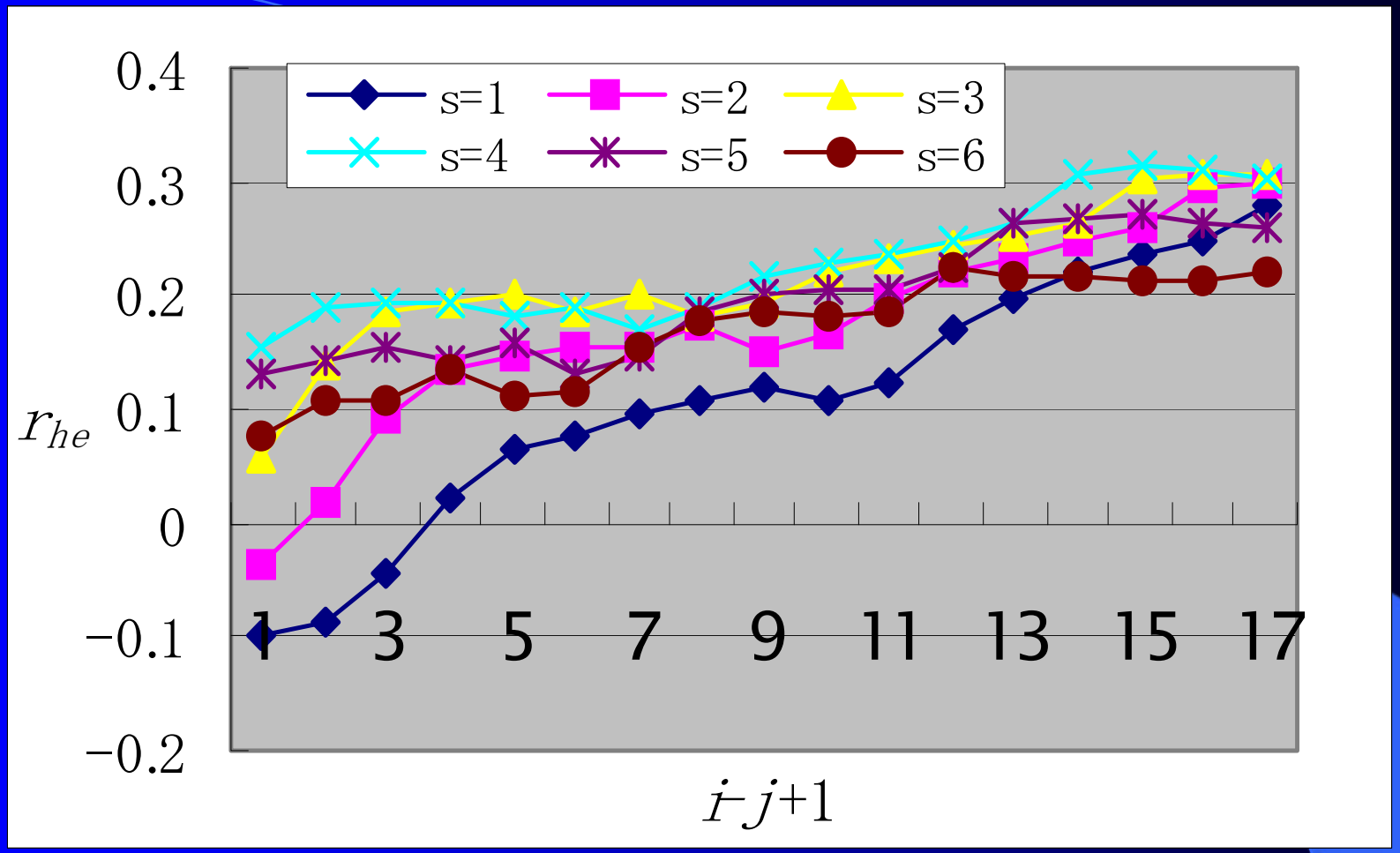
Algorithm for automated rule mining



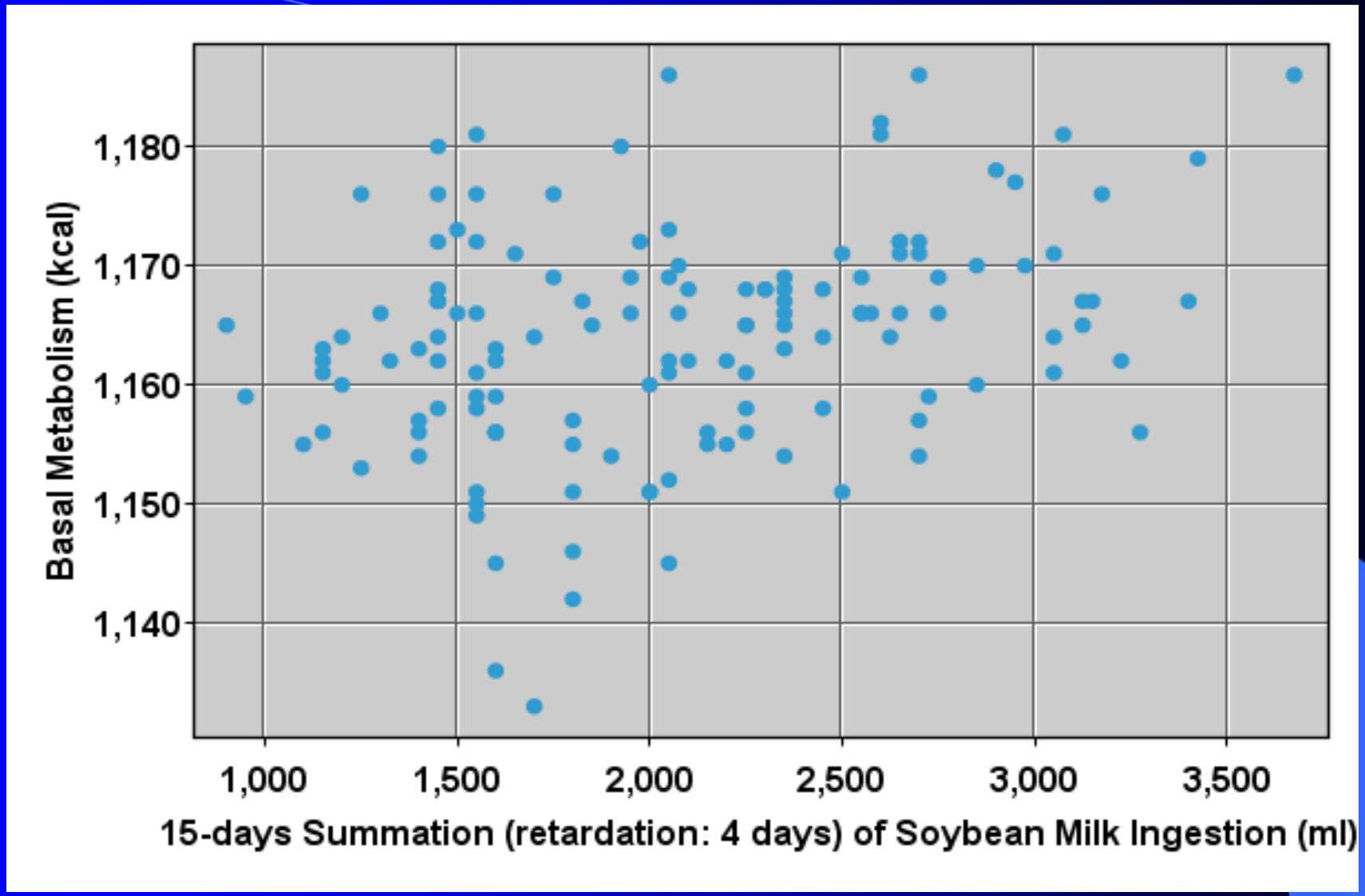
Correlation coefficient r_{he} between body-fat percentage and summation of soybean milk ingestion as a function of summation-day number (*i-j+1*) for various values of retardation parameter *s*



Typical scatterplot of body-fat percentage (*n*:139, *max*:27.0, *min*:22.9, *med*:25.0, *SD*:0.906) versus soybean milk ingestion (correlation coefficient: -0.470)



Correlation coefficient r_{he} between basal metabolism and summation of soybean milk ingestion as a function of summation-day number (*i-j+1*) for various values of retardation parameter *s*



Typical scatterplot of basal metabolism (*n*:138, *max*:1186, *min*:1133, *med*:1160, *SD*:9.48) versus soybean milk ingestion (correlation coefficient: 0.313)

Conclusion

Time-series data analyses for healthcare-data-mining process in the personal dynamic healthcare system (PDHS) presented here produced interesting results concerning the relations between lifestyle data and health data for volunteer users of PDHS.

Example: Daily soybean protein ingestion gradually decreased body-fat percentage (12 days successive ingestion was most effective with 4 days delay), and gradually increased basal metabolism (14 days successive ingestion was most effective with 4 days delay).

References:

[1] Smyth P and Goodman RM. An information theoretic approach to rule induction from database. IEEE Trans. Knowledge and Data Engineering 1992: 4(4) 301-316.

[2] Takeuchi H, Hashiguchi T, and Shintani T. Personal dynamic healthcare system utilizing mobile phone and web technologies. Proc. 2nd Int. Conf. on Advances in Medical Signal and Information processing 2004: 304-307.

[3] Takeuchi H, Kodama N, Hashiguchi T, and Mitsui N. Healthcare data mining based on a personal dynamic healthcare system. Proc. 2nd Int. Conf. on Computational Intelligence in Medicine and Healthcare 2005: 37-43.

[4] Takeuchi H, Kodama N, Hashiguchi T, and Hayashi D. Automated healthcare data mining based on a personal dynamic healthcare system. Proc. 28th IEEE EMBS Annual Int. Conf. 2006: 3604-3607.

Acknowledgments: We thank Dr. D. Hayashi of the University of Tokyo for his helpful discussions. This work was supported by the Grants-in-Aid for Scientific Research, Ministry of Education, Culture, Sports, Science and Technology.

Contact details: Hiroshi Takeuchi, Department of Healthcare Informatics, Takasaki University of Health and Welfare, Naka-Orui 37-1 Takasaki, Gunma, 370-33 Japan, Tel:+81-27-352-1290, Fax:+81-27-353-2055, Email: htakeuchi@takasaki-u.ac.jp