Prevention of Life-Style Related Diseases Using Machine Learning

Prof. Pallavi N[1], N L Subramanya Tejas[2]
[1, 2]Dept of Computer Science & Engineering
[1, 2]Atria Institute of Technology Bangalore, India

Abstract- *Diseases that are associated with the way a person or group of people live are known as lifestyle diseases. Lifestyle disease has been a sprawling issue not only in India, but worldwide. Recently, with the westernization of culture and eating habits, lifestyle-related diseases, such as cancer, heart disease, apoplexy, have become the main reason of people's death. The accumulation of improper lifestyles such as stress, eating habits, smoking, drinking, is the inducement of the lifestyle related disease. It is also found that lifestyle-related disease could be prevented with balanced lifestyles. Healthcare industry collects enormous disease-related data that is unfortunately not mined to discover hidden information that could be used for effective decision making. Therefore, this research focuses on predicting lifestyle diseases that an individual might be susceptible to. Moreover, we propose an intelligent low-cost alternative to detect possible genetic disorders caused by unhealthy lifestyles.*

Keywords- Fuzzy theory, Features, Machine Learning, Prediction sub system, Check sub system

I. INTRODUCTION

The chronic diseases induced by accumulation of improper lifestyles such as stress, low fruit and vegetable intake, smoking, alcohol use, lack of physical activity, etc. are called lifestyle-related diseases [1]. They are grave diseases, as it develops in the absence of subjective symptoms, such as pain,fever, etc., and by the time symptoms are manifested, quality of life is lowered, loss of life may ensue. Recently, it is also found that lifestyle-related disease can be prevented with balanced lifestyles, which concern with a mass of medical knowledge. If a publichealth nurse is always available, health check or healthadvice could be got whenever and wherever, so that everyone could contribute to the society by health andlong life. However, it is impossible in today's densely populated society.

Therefore, instead of the public health nurse, the construction of a health check and predictionsystem for lifestyle-related disease prevention is proposed.

Recently, based on bivalent logic and Internet technology, health-check systems have been developed.It is easy for people to know the possibility of lifestylerelateddisease by simply choosing Yes/No options of questions about symptoms and lifestyles. These researches play an important role in the health improvement. However, since only two stations (Yes/No) are considered, reasoning evidences are not sufficient. On the other hand, since fuzzy theory was proposed by L.A. Zadeh, many researches have made efforts on building of medical diagnosis supporting systems based on fuzzy reasoning [2-4],which have been applied widely in the special medicine field. However, those systems are impossible to be used by the common man. Therefore, in this paper, fuzzy theory is employed in building a health check and prediction system that caters to a larger audience.

II. SYSTEM DESIGN

The health check and prediction system consists of two parts, Check Sub System and Prediction Sub System, as shown in Fig.1. The Check Sub System is composed of Personal Constitution Characteristic Database (PCCDB), Medicine Knowledge Database(MKDB), and Reasoning Engine.

According to medical knowledge, current risk of person's condition is checked considering physiological data and subject symptoms. Prediction Sub System also consists of Advice Information Processing Module except MKDB and Reasoning Engine.

The Advice Information Processing Module calculates the future risk of person's condition according to medicalknowledge, lifestyle, hereditary and current risk. Furthermore, advice is also given for promoting health.

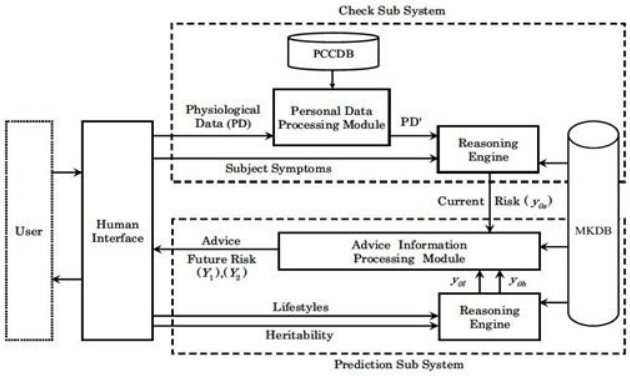

Fig.1 Outline of the health check and prediction system

III. METHOD ANALYSIS

Quantification and Formulation of Medical Knowledge

Since the accumulation of improper lifestyles is the inducement of the lifestyle-related disease, different from some other medical expert systems, besides " relation between the disease and the symptoms," "relation between the disease and the lifestyles" is considered when MKDB of the health check and prediction system is composed. Furthermore, "relation of disease" is also considered, so that the hereditary determinant could be related to the possibility of the disease. Since it is too hard to control the external environment factors, such as the pathogen, poisonous substances, and accidents intentions, they are not considered in this paper. The medicalknowledge concerning with hypertension is shown inTable.1 as an example. Since more detailed check results could be acquired by considering the degree of a symptom besides Yes/No, fuzzy variables are used to quantify the degrees of the medical knowledge. Fuzzy variables, "None (N)," "Mild (M)," "Slightly Serious (SS),""Serious (S)," and "Very Serious (VS)" are used to describe the degree of the symptoms and lifestyles. For the risk of diseases, fuzzy variables of "Almost zero(Z)," "Slight (S)," "Slightly high (SH)," "High (H),"and "Very high (VH)" are used. Then the MKDB is built by the formulated fuzzy rules shown as (1). xSi, xLi and xHi are defined as items of symptom, lifestyle, and heritability of people respectively. dSi, dLi and dHi are defined as the degree of each item. Let RS, RL and RH be the risk of the lifestyle-related disease concerning with each precondition.

Table.1 Medical Knowledge of Hypertension

Symptom		Lifestyle	
x_{s1}	Blood pressur is high	x_{L1}	Salinity excessive
x_{s2}	Dizziness	x_{L2}	Being over weight
x_{s3}	Shoulders are stiff	x_{L3}	Not exercising
x_{s4}	Irregular or rapid heartbeat	x_{L4}	Stress
x_{s5}	Eyesight falls	x_{L5}	Insomnia
x_{s6}	Headache	x_{L6}	Bibulosity
x_{s7}	Lassitude	Heritability	
x_{s8}	Numbness of hands or feet	x_{H1}	Some relatives suffers from hypertension
x_{s9}	Facial flushing		
x_{s10}	A buzzing noise in the ear	x_{H2}	Diabetic or degeneration of organ or cardiovascular disease

If {xSi is dSi} then RS.
If {xLi is dLi} then RL. (1)
If {xHi is dHi} then RH.

According to Table.1, fuzzy rules of hypertension based on subjective symptoms were formulated as follows.

$$\begin{array}{llll} \textit{If } x_{s1} \textit{ is VH,} & \textit{If } x_{s2} \textit{ is VH,} & \ldots, \textit{ If } x_{s10} \textit{ is VH} & \textit{then } R_s \textit{ is VH} \\ \vdots & \vdots & \vdots & \vdots \\ \textit{If } x_{s1} \textit{ is N,} & \textit{If } x_{s2} \textit{ is S,} & \ldots, \textit{ If } x_{s10} \textit{ is H} & \textit{then } R_s \textit{ is H} \\ \vdots & \vdots & \vdots & \vdots \\ \textit{If } x_{s1} \textit{ is N,} & \textit{If } x_{s2} \textit{ is N,} & \ldots, \textit{ If } x_{s10} \textit{ is N} & \textit{then } R_s \textit{ is Z} \end{array}$$

Fuzzy rules concerning with lifestyles and heritabilityformulated by the same mechanism are omitted in this paper.

Distance-Type Fuzzy Reasoning Method

Based on the distance information between fuzzysets, the Distance-Type Fuzzy Reasoning (DTFR)Method is built [5]. Due to DTFR Method satisfies the decomposition rule and the convex and asymptotic characteristic of the reasoning result has been proved , DTFR Method could easily respond to the fact of deleting old knowledge or adding new knowledge and cases in the medicine database. Hence, a reasoning algorithm of the check and control of physiological information is built up by using DTFR Method.

As in [5] — according to the fuzzy reasoning rules of m inputs x1 ~ xm and 1 output y, as shown in (2), the DTFR Method and its features are discussed. Antecedents, consequents, given facts and reasoning result is represented by Aij, Bi, Aj, and B respectively, for i=1 to n, j=1 to m.

$$\begin{array}{lll} R^i & : x_1 = A^{i1},\ x_2 = A^{i2}, & , \ x_m = A^{im} => y \ \ B^i \\ \text{Fact} & : x_1 = A^1,\ x_2 = A^2, & , \ x_m = A^m \\ \hline \text{Result}: & & y \ \ B \end{array} \qquad (2)$$

DTFR Method involves the following three steps:

Step 1: Calculate the distance d_{ij} (A^{ij}, A^j) between the antecedent A^{ij} and A^j by the method given in [4], and then calculate d_1 to d_n using (3).

$$d_i = \sum_{j=1}^{m} d_{ij}(A^{ij}, A^j) \qquad (3)$$

Step 2: Calculate the α-level set of reasoning result B as follows. Sup(B_α) is an upper limit of crisp subsetB_α, and inf(B_α) is a lower limit of crisp subset B_α.

$$B_\alpha = [\inf(B_\alpha), \sup(B_\alpha)] \qquad (4)$$

$$\inf(B_\alpha) = \frac{\sum_{i=1}^{n}\left[\inf\left(B_\alpha^i\right)\prod_{j=1,j\neq i}^{n} d_j\right]}{\sum_{i=1}^{n}\prod_{j=1,j\neq i}^{n} d_j} \qquad (5)$$

$$\sup(B_\alpha) = \frac{\sum_{i=1}^{n}\left[\sup\left(B_\alpha^i\right)\prod_{j=1,j\neq i}^{n} d_j\right]}{\sum_{i=1}^{n}\prod_{j=1,j\neq i}^{n} d_j} \qquad (6)$$

To avoid confusion, sup(B_α)and inf(B_α) also can be represented by $\overline{B\alpha}$ and $\underline{B\alpha}$ respectively. When interval is considered (4) change to (7).

$$B_\alpha = \frac{\sum_{i=1}^{n}\left[B_\alpha^i \prod_{j=1,j\neq i}^{n} d_j\right]}{\sum_{i=1}^{n}\prod_{j=1,j\neq i}^{n} d_j} = \left[\frac{\sum_{i=1}^{n}\left[\underline{B_\alpha^i}\prod_{j=1,j\neq i}^{n} d_j\right]}{\sum_{i=1}^{n}\prod_{j=1,j\neq i}^{n} d_j}, \frac{\sum_{i=1}^{n}\left[\overline{B_\alpha^i}\prod_{j=1,j\neq i}^{n} d_j\right]}{\sum_{i=1}^{n}\prod_{j=1,j\neq i}^{n} d_j}\right] \qquad (7)$$

Step 3: Obtain the reasoning result B as (8) according to the composition theorem (also called the decomposition theorem) [5].

$$B = \bigcup_\alpha \alpha . B_\alpha \qquad (8)$$

IV. PREDICTION OF LIFESTYLE-RELATED DISEASE

Since lifestyle-related disease concerns withperson's lifestyles, it cannot be said that future risk oflifestyle-related disease is zero even though the currentcondition is quite well. In other words, it is importantto sound the alert to people according to the predictionof future risk of lifestyle-related disease. Therefore, aprediction method is proposed. As former explanation,lifestyle-related disease is induced by accumulation ofimproper lifestyles. On the contrary, the risk oflifestyle-related disease will reduce by improvingcurrent lifestyles. In this paper, the two kinds ofphenomenon discussed above are quantified by (9) and(10) respectively. The reasoning results of subjectsymptoms, lifestyles and heritabilities are representedby y_{0s}, y_{0l}, and y_{0h} respectively. k1 is an adjustableparameter which represents the importance ofinfluence from lifestyles and heritabilities. The largerk1 is, he influences grows. k2 represents the importance of the influence comes from lifestyles compared with that from heritabilities.

$$Y_1 = y_{0s} + \frac{k_1}{1+k_2}(k_2 y_{0l} + y_{0h})(1 - y_{0s}) \qquad (9)$$

$$Y_2 = y_{0s} - \frac{k_1}{1+k_2}(k_2 y_{0l} + y_{0h}) y_{0s} \qquad (10)$$

V. CONCLUSION

In this design of life-style disease predictionsystem, the fuzzy set theory was employed to deal withthe fuzziness of degrees of the subject symptoms,lifestyles and heritabilities. Therefore, a more reasonableprediction result could be obtained. Moreover, since the system was built based on Distance-Type FuzzyReasoning Method, updating of the medicalknowledge database became easy and quick.

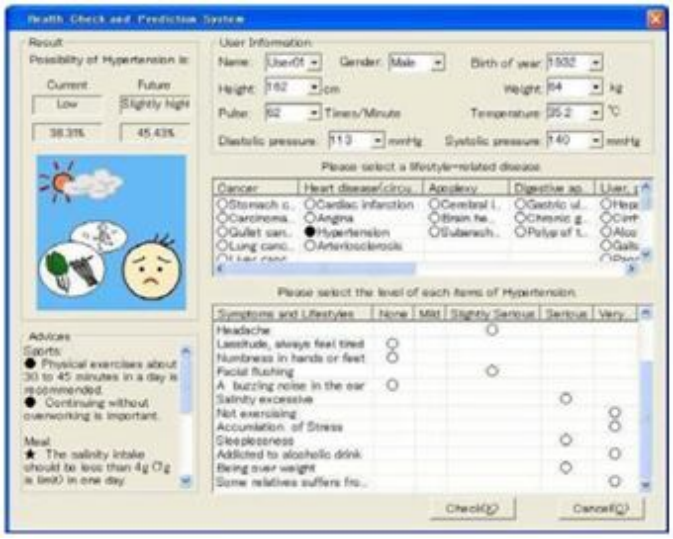

Fig.2 Human interface of the system

Furthermore, prediction of lifestyle-related disease waspproposed as while as the condition check facing toeveryman. Lastly, the usability of proposed methodwas discussed by experiments of developed system.Further clinical test will be carried out in the future toverify the effectiveness of proposed methods anddeveloped system.

REFERENCES

[1] Kagawa, Y., *Preventing from Lifestyle-Related Disease*, Iwanami Shoten, Tokyo, 2000.

[2] Sharma, M. and Majumdar, P.K., 2009. Occupational lifestyle diseases: An emerging issue. *Indian Journal of Occupational andEnvironmental medicine*, 13(3), pp. 109–112.

[3] Nguyen Hoang Phuong, Hung H. Dang, Nadipuram R. Prasad, "Developing a Supporting Expert System for Lung-Disease Diagnosis Using Fuzzy Logic", *Biomedical SoftComputing and Human Sciences*, Vol.5, No.2, 2000, pp. 37-44.

[4] Hisakazu Ogura, Hidetoshi Ukai, Haruhiko Shirai, Junji Nishino, Tomohiro Odaka and Shuzo Oshita, "Multistage Fuzzy Inference for Blood Pressure Control in Anesthesia and Knowledge Extraction from Time-Series Clinical Data by GA", *Journal of Japan Society for Fuzzy Theory andSystems*, Vol.14, No.2, 2002, pp. 228-239. (in Japanese).

[5] Pattekari, S.A. and Parveen, A., 2012. Prediction system for heartdisease using Naïve Bayes. *International Journal of AdvancedComputer and Mathematical Sciences*, 3(3), pp. 290–294.