TDR-S: A 2-cycle Health Condition Classification System

1. OVERVIEW

TDR-S is a 2-cycle health condition classification system. It consists of a fast answer cycle and a slow answer cycle. The fast answer cycle offers immediate answer using an outliner based detection model that takes into account both F_c and F_s. On the other hand, the slow answer cycle is a social network based model that allows beginners to learn from experts, who are ususally minorities in the social network, in the social network. The slow answer cycle offers answer when a person that satisfies certain requirement offer answer. More details are presented in Section 2 and Section 3.

2. FAST CYCLE: OUTLINER DETECTION MODEL

In this section, we introduce a classification model that takes into account both F_c and F_s.

2.1 Related Features

2.1.1 *Scientific Measurements F_s:*

Blood pressure, SPO2, EKG, and body temperature are standard scientific features that can be easily measured by sensors. Usually, they can be accurately measured. Also, there are also standard value boundaries for determining abnormal values.

2.1.2 *Chinese Medicine Measurements F_c:*

Tongue, fatigue, weak breadth, pulse, sweaty are 5 traditional Chinese medicine measurements that can be felt or sensed by people themselves or determined by Chinese medicine physicians. However, the values felt or sensed by people themselves may sometimes be subjective.

2.1.3 *Personal Information F_p:*

Sex, age, and race are useful personal information that can be use to distinguish a person's health reasonable health condition since people with different sex, age, and race usually have different normal values for F_s and F_c.

2.2 Methodology

2.2.1 *Outliner Detection for F_c*

Firstly, all people are separated into group based on their F_p. Secondly, all data points that are within the same group are put into a multi-dimensional space based on F_c. The center point c of these data point is the mean of all data points. We regard all data points that are τ from the center point as outliners and the other $p\%$ as normal data points. τ is a parameter that can be determined based on related medical literatures.

$$M_c = \begin{cases} 1, & \text{if } dist(F_c, c) \geq \tau. \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

,where 1 indicate abnormal and 0 indicates normal.

In Figure 2, people with $F_p = \{M, 20-30, Asian\}$ are put into the same group and p is set to 95%. Hence, all red dots are regarded as outliners.

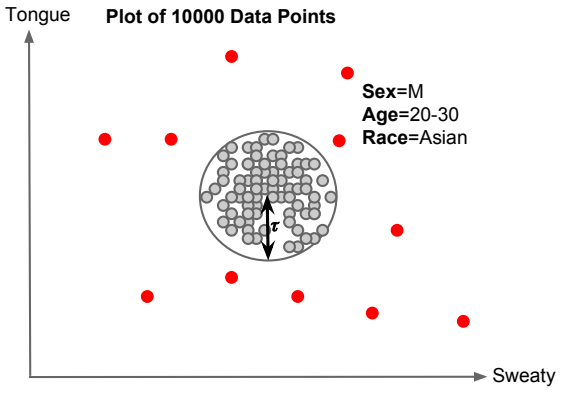

Figure 1: Outliner Detection

2.2.2 *Model using F_c and F_s*

M is a model for fast cycle. We regard every parameter in F_s as a M_i^s. M_i^s is 1 if the value is an abnormal value based on F_p of that person; otherwise, M_i^s is 0.

$$M_i^s = \begin{cases} 1, & \text{if } F_s \text{ is abnormal, given } F_p. \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

$$M = max(M_c, M_1^s, M_2^s, M_3^s, M_4^s)$$

Parameters of M_c are determined based on feeling and experience of people. Therefore, it is possible that error may occur. M_i^s and abnormal boundaries of them are scientific measurements. Therefore, even M_c is 0, if any M_i^c is 1, we should still regard M as 1 (abnormal).

3. SLOW CYCLE: SOCIAL NETWORK MODEL

This section will first introduce the basic framework of Social Network Model and how it is integrated into the TDR system. And then describe the two track of the Social Network Modeling: (1) how to upload the judgements to the database and (2) how to update the weight of each profile in the model.

3.1 Social Network Model Framework

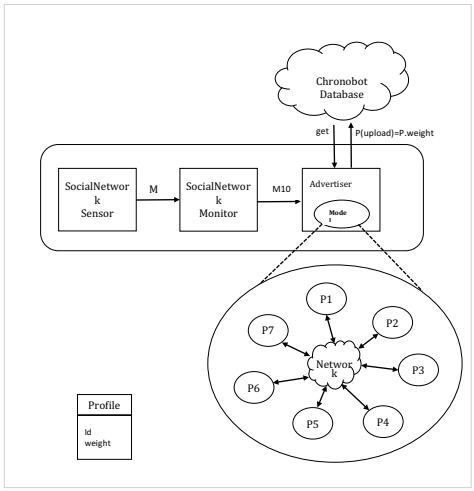

Figure 2: SocialNework Framework

3.2 Teacher-Student Model

A profile in the group with higher weight is more reliable in judging others and we call this kind of profile teacher profiles. In the group model, there is one or more profile which is called teacher profile who has a higher weight while the other profiles own lower weights. The basic mode for each profile are two attributes: a $profile_id$ which can uniquely identify the profile in the model and a $profile_weight$ which represents the reliability of the judgements generate by the profile.

3.2.1 Judgement Upload Track

Judgements from different profile int the model are not treated equal in our model. A teacher profile may always make judgements corresponding with the real status of the target. In other words, a teacher is more likely to generate the correct judgements in the model. Thus we should accept all the judgements from the teacher and treat them as Oracle who always knows the truth. Accordingly, the weight of a teacher is set to be 1. On the other hand, the tiro-students who don't know how to make Chi judgements are less likely to generate reliable judgements and thus their judgements are not accepted by the system. In fact, there are intermediate students whose judgement ability is between a tiro-student and the teacher and there must be a threshold θ where the weights large than it are treated as teacher-like weights. Thus we should find out where the θ stays and distinguish the teacher-like students and accept their judgements. As time goes by, students are going to behave like teacher and generate reliable judgements.

3.2.2 Judgement Upload Track

The weight of a profile is related to the average distance of judgements with the teacher profile in history. For each time point n where the computation cycle happens, the weight of a $profile_i$ can be updated by the following equation.

$$W_i = \frac{\sum_{t=1}^{n} dist(J_i t - J_o t)}{n} \quad (3)$$

, where $J_i t$ is the judgement at time t by $profile_i$ and $J_o t$ is the judgement by the teacher at time t. The W_i considers all the history distance of a specific profile to the teacher. If a student knows nothing about making Chi judgement at first and finally grasp the method, and his judgments will be accepted after a time point t.

4. EXPERIMENT

4.1 Experiment Design

4.1.1 Fast Cycle

In this experiment, we want to evaluate the effectiveness of our outliner detection model in Section 2.2.1. We want to show that:

1. most people are healthy people and they would have similar F_c, and
2. most people can approximately determine their F_c.

Since we are unable to obtain real F_c data, in this experiment, we will use similar data to evaluate the effectiveness of our proposed model. We perform our experiment in Amazon Mechanical Turk. Given 5 pictures of 5 different faces, we ask 1000 workers to determine the size of nose, eyes, and mouth in every picture. For all 5 pictures, persons in the pictures have similar size of nose, eyes, and mouth. Worker can offer a score from 1 to 5, where 1 is the smallest size and 5 is the biggest size. Then, we put all data points together and build a model using method in Section 2.2.1. We want to show that it is possible to find a dense region in all data points, which implies most people would have similar sense while a small number of people are different from majority and people are able to offer approximate score for things that they understand.

4.1.2 Slow Cycle

To Prof. Liang, could you please fill up this subsection? Since no real data, in order to make a good social network learning model, we design a similar experiment named gray-scale learning. (1)Definiton: two kinds of person, one is master(answers always is right), another are students(keep learning for improving judgement). (2)Basic knowledge: gray-scale has 256 value, from 0 to 255. Before the expereiemnt, all students are given a pitcure with 5 five colors which show five different gray values. (3)learning process: In every learning cycle, the master and students are given a new picture which contains five random colors. Students need to fill out gray values for every colors based on their basic knowledge and knoelwedge learnt in previous cycles. The master also has to fill out gray values for all five random colors. The answer offered by the master are given to all students at the end of this cycle so that students can learn from the master's answer, which is always treated as the right answer. After got the right answer from the master, they will enter next slow cycle. (4)We will record every judgement result for every students and the master, and the records will be used for evaluating the effectiveness of the social network model.