

진단알고리즘구성에서 증상들의 수값화와 증상들사이 거리계산의 한가지 방법

송원일, 김택중

질병진단프로그램들에서 쓰이는 진단알고리즘은 환자가 입력한 여러가지 증상들을 분석종합하여 해당하는 진단결과를 예측하도록 설계한다.

일부 선행한 진단지원체계들[2]에서는 의학적증상자료들을 그대로 리용하는것으로 하여 리용에서 일정한 제한성을 가지게 되므로 종합적인 질병들이 아니라 제한된 부분의 질병들에 대한 진단을 진행하였다.

실례를 들어 유선암진단, 심장질병진단, 장티브스성열병진단 및 치료 등과 같이 전반적인 질병진단이 아니라 일정한 분야의 질병진단이었다.

한편 증상자료를 수값화하기 위한 선행연구[1]에서는 증상들이 해당하는 질병들에 들어 있으면 1로, 없으면 0으로 표시하여 1로 표시되는 개수가 제일 많은 질병을 진단하려는 질병의 후보로 선정하는 방법을 리용하였다. 이 방법은 몇개의 질병들에 대하여 증상들이 개수가 많지 않은 경우에는 효과적이지만 진단하려는 질병들의 개수가 수천개 지어는 만여개를 넘게 되면 이 방법은 적합치 않다. 특히는 질병들마다 증상들의 개수가 서로 다른 경우에 위의 방법은 전혀 적용이 불가능하게 된다.

게다가 증상들사이의 유사성과 관련하여 언급하지 않았다.

론문에서는 보다 일반화되고 종합적인 진단알고리즘을 구성하는데서 중요한 의의를 가지는 증상들의 수값화와 증상들사이의 거리를 계산하는 한가지 방법을 고찰하였다.

1. 증상들의 수값화(증상특성값)

증상들의 수값화는 증상들을 요소별로 분할하고 매 요소들이 담고있는 내용을 수값으로 표시하는것으로서 증상들의 내용구성방식을 리용하여 규정하였다.

증상들의 내용구성은 대체로 어느 부위에서, 무엇이 어떤가, 어느 정도로, 언제, 어떤지속인가, 어느 방향으로 이동하는가 등의 요소들로 이루어진다고 볼수 있다.

즉 증상 x 를 다음과 같이 형식화할수 있다.

$$x = \{c_1, c_2, \dots, c_p\}$$

여기서 c_i 는 요소, p 는 요소들의 개수이다.

증상의 수값화는 증상의 매 요소들에 요소값들을 결정하고 그에 기초하여 증상의 특성값을 계산하는 방법으로 진행한다.

1) 요소값의 결정

어떤 증상 x_k 에 대하여 이 증상의 i 번째 요소 c_i 의 요소값 $ch_i(i=\overline{1,p})$ 는 해당 c_i 요소에 들어있는 내용들이 많고 적음에 따라 1~4자리의 옹근수값을 일의적으로 할당하는

방법으로 결정하였다.

i 째 요소 c_i 의 자리수를 $t_i, i=\overline{1, p}$ 로 표시하면 해당 증상에 대한 전체 자리수는 $\sum_{i=1}^p t_i$ 로 된다.

2) 증상의 특성값계산

어떤 증상 x_k 의 특성값 $ch(x_k)$ 는 매 요소들의 요소값들을 결합한 값으로서

$$ch(x_k) = \sum_{i=1}^p ch_i^{(k)} \times 10^{r_i}, \quad k=\overline{1, q} \quad (1)$$

로 쓸수 있다. 여기서

$$r_i = \sum_{j=i+1}^p t_j$$

이고 q 는 취급되는 증상들의 총개수이다.

실례로 《심한 머리아픔이 오래 지속된다.》라는 어떤 증상 x_k 를 놓고 수값화를 보기로 하자.

이 증상은 《어느 부위에서》(머리), 《무엇이 어떤가》(아프다), 《어느 정도로》(심하다), 《어떤 지속인가》(오래 지속)의 4개의 요소로 구성되어있다.

앞에서 증상들의 모든 요소값들을 일의적으로 결정하였다고 언급하였다.

만일 《머리》에 해당하는 요소값 $ch_1^{(k)} = 100$, 《아프다》에 해당하는 요소값 $ch_2^{(k)} = 002$, 《심하다》에 해당하는 요소값 $ch_3^{(k)} = 3$, 《오래 지속》에 해당하는 요소값 $ch_4^{(k)} = 4$ 라고 할당되어있다면 이 증상의 특성값은 $ch(x_k) = 10000234$ 로 된다.

다음의 표에 일부 증상들과 그것들의 특성값들을 보여주었다.

표. 몇가지 증상들과 특성값

번호	증 상	특성 값
1	심한 머리아픔이 오래 지속된다.	10000234
2	삼키기장애가 짧은 기간에 심해진다.	20000500
3	가슴아픔이 갑자기 심하게 나타난다.	30001101
4	손발에서 심한 저림증상이 있다.	60040302
5	식후에 소화장애가 있다.	50030400

이와 같은 방법으로 모든 증상들을 다 수값화하였다.

2. 증상들사이의 거리계산

증상들사이의 거리는 앞에서 언급한 증상의 특성값들을 리용하여 구할수 있다. 그러한 의미에서 증상들의 특성값계산은 증상들사이 거리계산의 전제조건이라고 볼수 있다.

어떤 두 증상 x_i 와 x_j 사이의 거리를 구하는 문제를 보기로 하자.

① 요소들사이의 거리

먼저 증상들의 요소들사이 거리를 구한다.

두 요소들사이의 거리는 해당 요소들이 증상적으로 서로 련관이 없을 때에는 어떤 고정된 정수값으로 주고 련관이 있을 때에는 련관정도에 따라 일정한 한계범위내에서의 정수값을 취하게 한다.

실례를 들어 부위요소를 보면 《배》부위와 《다리》부위는 련관이 없는것으로 보고 《다리》부위와 《발》부위는 련관이 있는것으로 본다.

정의 1 두 증상 x_i 의 요소 $c_i^{(k)}$ 와 x_j 의 요소 $c_j^{(k)}$ ($k=1, \overline{p}$) 사이의 거리는

$$d_k(x_i, x_j) = \begin{cases} d \max_k (d \max_k > 0), & i \neq j, \quad \text{련관이 없을 때} \\ 0, & i = j \\ d_{ij} (d_{ij} > 0), & i \neq j, \quad d \min_k \leq d_{ij} < d \max_k, \text{련관이 있을 때} \end{cases} \quad (2)$$

이다. 여기서 $d \max_k$ 는 두 증상들의 k 째 요소들사이 거리의 윗한계를, $d \min_k$ 는 두 증상들의 k 째 요소들사이 거리의 아래한계를 규정해주는 값이며

$$d \min_k > d \max_{k-1} \quad (k=2, \overline{p})$$

이다.

② 증상들사이의 거리

요소들사이의 거리들을 구한데 기초하여 두 증상 x_i 와 x_j 사이의 거리를 다음과 같이 정의할수 있다.

정의 2 두 증상 $x_i, x_j \in X$ 의 거리는

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^p (d_k(x_i, x_j))^2} \quad (3)$$

이다.

정의 1과 2에서 정의된 요소들사이, 증상들사이의 거리들이 거리의 공리를 만족시킨다는것은 분명하다.

맺는 말

진단프로그램들에서 리용하는 모든 증상들을 수값화하고 임의의 두 증상들사이의 거리를 계산함으로써 증상들의 류사한 정도를 량적으로 평가할수 있는 수단을 가지게 되었다.

따라서 환자들이 입력한 증상들을 놓고 여러 질병들의 증상들과의 류사성에 기초하여 진단을 하는 알고리즘을 구성할수 있게 하였다.

결국 환자가 입력한 증상들이 어느 질병의 증상렬에 많이 들어있는가에 따라 질병을 진단하는 선행한 진단수법들에 비하여 보다 정확한 진단결과를 내릴수 있는 방법을 확립하였다.

또한 의학적화상자료나 기타 림상자료들을 그대로 리용하여 진단을 하는 선행한 진단수법들에 비하여 특정한 1~2개 분야가 아닌 전반적이며 종합적인 분야들에 대한 질병 진단을 진행할수 있는 방법을 확립하였다.

참 고 문 헌

- [1] 백형남 등; 정보과학, 3, 27, 주체98(2009).
- [2] Chukwuedozie N. Ezema et al.; International Journal of Engineering Research and Reviews, 4, 1, 82, 2016.

주체106(2017)년 11월 5일 원고접수

A Method for the Numerical Value of the Symptoms and Estimating the Similarity Degree between Symptoms in the Composition of the Diagnostic Algorithm

Song Won Il, Kim ThaeK Jong

In this paper, by making the numerical value of symptoms and calculating the distance between every symptom in the diagnosis support systems, analogy degree of the symptoms can be estimated in a quantitative way.

So, it became possible to make an algorithm that makes diagnoses based on the similarity between symptoms of diseases.

Key words: numerical value of symptoms, distance, diagnostic algorithm