

한의학자료의 수량화에 대한 연구

신 양규¹

요약

한의학에서 환자의 병적상태를 분석하고 진단하기 위한 전문가시스템은 진단과정에서 환자로부터 받아들인 지식에 기초하므로 지식은 정확하게 표현되어야 한다. 그러나 대부분의 한의진단과정에서 환자로부터 받아들인 지식은 주로 질적자료에 기초하여 기준이 애매한 개념의 지식을 많이 포함하므로 이에 대한 처리가 요구된다. 본 논문에서는 한의전문가시스템 구축에 이용되는 정보의 정확도를 높이기 위하여 Hayashi의 수량화II류를 기반으로 하여 한의진단과정에서 얻어지는 지식을 수량화하는 방법을 제안하였다.

주제어: 수량화 방법, 질적자료, 한의진단지식.

1. 서론

한의학에서 증과 변증의 현대 연구(양사두와 진귀연, 1991)라는 것은 한의학에서 지금까지 사용되어오던 방법 이외에 다른 방법론들을 도입하여 변증과 증후를 해석하고 표준화, 객관화, 응용범위의 확대 등을 도입하고자 하는 것이다. 본 연구에서는 변증과 증후를 표준화, 객관화하기에 적합한 통계적 방법에 대하여 연구해보고자 한다.

증후의 입장에서 보면 현재까지 변증의 지표가 되어온 증후는 대개 사진을 통해서 얻을 수 있는 것들이었으며 각 변증체계의 이론과 맞물려 그 변증체계에서 유용하고 해석 가능한 증후들을 위주로 각 증의 증후요건이 정해졌다. 그런데 증후가 각 증에 대해 특이성을 가질수록 증의 판별은 용이하여진다. 그러므로 각 변증체계내의 증후의 변증에 대한 의의를 객관적으로 판정하는 방법론이 필요하다.

증후의 변증에 대한 의의 판정에 이용될 수 있는 통계적 방법론은 정준상관분석이나 정준판별분석을 생각 할 수 있으나 질적자료의 성질을 주로 가지는 한의학 자료의 특성에 합당하지 못하여 한의학적 정보를 제대로 도출할 수 없으므로 본 연구에서는 Hayashi박사(하야시 치키오)에 의해 개발된 통계분석기법인 수량화기법(Quantification method)를 이용하

¹(712-240) 경상북도 경산시 점촌동 산 75, 경산대학교 자연과학대학 통계학과 조교수.

여 진단지식을 객관화 할 수 있는 방법에 대하여 연구해 보고자한다. 수량화된 진단지식을 통하여 변증지표로서의 증후를 최대한 정확하고 효율적으로 해석함으로써 증후의 정체가 명확해져 변증진단의 표준화와 객관화가 이루어질 수 있으며 이를 통하여 증후모형을 만드는 기반을 제공할 것이다. 나아가 Shin(1997)에 의하여 제안된 CLP(R)언어를 이용한 한의진단모델의 추론과정에서 발생하는 불확실한 진단지식의 처리방법과 병행하여 한의진문가시스템 구축을 위한 객관화된 지식습득 방법으로 사용될 수 있으리라 기대된다.

본 논문에서는 수량화 II류(Theory of Quantification II)를 기반으로 하여 한의진단과정에서 얻어지는 진단지식 즉 증후와 증형을 수량화하는 방법을 제안하였다. 제 2절에서는 본 연구에 필요한 수량화 기법에 대하여 살펴보았고(허명희,1992; 노형진,1990) 제 3절에서는 한의진단과정에서 얻어지는 진단지식 즉 증후와 증형을 수량화하는 방법을 제안하였고, 마지막에 예제로 중풍환자의 진단지식을 수량화하였다.

2. 수량화방법

Hayachi박사에 의해 제안된 수량화기법은 질적인 자료를 분석의 목적에 맞도록 최적인 수량이나 점수를 주어 양적인 자료로 변환한 다음 다변량적인 해석을 행하는 분석방법이다(허명희,1992; 노형진,1990). 수량화의 방법을 크게 나누면 외적 기준이 있는 경우와 외적기준이 없는 경우로 분류된다. 여기에서 외적기준이란 설명하고 싶거나 혹은 예측하고 싶다고 생각하고 있는 것으로 외적기준이 있는 경우의 수량화법에는 수량화 I류, II류등이 있고 외적기준이 없는 경우의 수량화법에는 하야시의 수량화 III류, IV류, 최소차원해석(MDA), 다차원척도법(MDS)등이 있다.

본 연구를 위하여 수량화 I류, II류, III류, IV류에 대하여 간단히 설명을 하고자 한다.(1,4)

수량화 I류는 질적인 요인에 관한 정보에 의거하여 양적으로 측정된 외적기준(목적변수)의 값을 설명 혹은 예측하기 위한 방법으로 수리적으로는 가변수에 대한 중회귀분석이라 볼 수 있다.

수량화 II류는 질적인 형태로 주어진 외적 기준을 질적인 요인에 의하여 예측 혹은 판별하는 방법으로 각 요인이 정량적으로 측정될 때는 판별분석을 적용할 수 있다. 수량화 III류는 개체의 여러 가지 범주로의 반응방식에 의하여 개체와 범주 양쪽을 수량화하고, 또한 그 수량을 이용하여 분류를 행하고자 하는 방법이다.

수량화 IV류는 대상들에 대하여 2개씩 짝을 지어 그 사이의 친근성을 정의하여 이 친근성에 의하여 대상에 수량을 부여하여, 친근성이 큰 쌍은 가깝게, 친근성이 작은 쌍은 멀리하도록 유클리드 공간 내에 배치하고자 하는 방법이다.

3. 한의진단지식의 수량화

한의진단모델에 의하면 진단과정에서 얻어지는 진단지식은 증형과 증후로 분류할 수

있다.(서산영웅,1980; 양사두와 진귀연,1991)

n 명의 환자들에 대하여 증형의 I 개 범주 가운데 i 번째 범주로 판정된 $n(i)$ 명의 환자를 모아 제 i 군이라 하고 제 i 군의 α 번째 환자가 R 개의 증후에 대하여 진단후 각 증후의 어느 범주에 해당하는가를 나타내기 위하여 다음과 같은 변수를 생각한다.

$$\delta_{i\alpha}(jk) = \begin{cases} 1 & i\text{번째 증형의 } \alpha \text{ 번째 환자가} \\ & \text{증후 } j \text{의 } k\text{번째 범주로 진단될 때} \\ 0 & \text{그 밖의 경우} \end{cases}$$

i 군의 α 번째 환자의 증형에 대한 양적인 표현을 $Y_{i\alpha}$ 라 두면 위의 가변수를 이용하여 다음과 같이 정의할 수 있다. 이때 c_j 는 각 증후의 범주를 나타낸다.

$$Y_{i\alpha} = \sum_{j=1}^R \sum_{k=1}^{c_j} a_{jk} \delta_{i\alpha}(jk)$$

이때 a_{jk} 는 증후 j 의 k 번째 범주에 부여하는 가중치로 환자의 증형을 가장 명확하게 진단하기 위하여 k 개의 증형간의 변동이 전체변동에 대하여 상대적으로 최대가 되도록 a_{jk} 를 정하고자 한다. 전체변동은 j 증후의 k 범주로 진단된 환자의 수를 $n(jk)$ 라 두면

$$\begin{aligned} SST &= \sum_{i=1}^k \sum_{\alpha=1}^{n_i} (Y_{i\alpha} - \bar{Y}_{..})^2 \\ &= \sum_{i=1}^k \sum_{\alpha=1}^{n_i} \left\{ \sum_{j=1}^R \sum_{k=1}^{c_j} a_{jk} \delta_{i\alpha}(jk) - \sum_{j=1}^R \sum_{k=1}^{c_j} \frac{n(jk)}{n} a_{jk} \right\}^2 \end{aligned}$$

이고, 증형간 변동은 i 증형에서 j 증후의 k 범주로 진단된 환자의 수를 $n(ijk)$ 라 두면

$$\begin{aligned} SSB &= \sum_{i=1}^k n(i) (\bar{Y}_i - \bar{Y}_{..})^2 \\ &= \sum_{i=1}^k n(i) \left\{ \sum_{j=1}^R \sum_{k=1}^{c_j} \frac{n(ijk)}{n(i)} a_{jk} - \sum_{j=1}^R \sum_{k=1}^{c_j} \frac{n(jk)}{n} a_{jk} \right\}^2 \end{aligned}$$

이다.

위의 식에서 j 증후의 k 범주와 j' 증후의 k' 범주로 진단된 환자의 수를 $n(jk, j'k')$ 라 두면

$$\begin{aligned} SST &= \sum_{j=1}^R \sum_{k=1}^{c_j} \sum_{j'=1}^R \sum_{k'=1}^{c_{j'}} \left\{ n(jk, j'k') - \frac{n(jk)n(j'k')}{n} \right\} a_{jk} a_{j'k'} \\ &= \sum_{j=1}^R \sum_{k=1}^{c_j} \sum_{j'=1}^R \sum_{k'=1}^{c_{j'}} f(jk, j'k') a_{jk} a_{j'k'} \end{aligned}$$

$$\begin{aligned}
 SSB &= \sum_{j=1}^R \sum_{k=1}^{c_j} \sum_{j'=1}^R \sum_{k'=1}^{c_{j'}} \left\{ \sum_{i=1}^k \frac{n(ijk)n(i'j'k')}{n(i)} - \frac{n(jk)n(j'k')}{n} \right\} a_{jk} a_{j'k'} \\
 &= \sum_{j=1}^R \sum_{k=1}^{c_j} \sum_{j'=1}^R \sum_{k'=1}^{c_{j'}} g(jk, j'k') a_{jk} a_{j'k'}
 \end{aligned}$$

으로 나타내진다. 이때

$$\begin{aligned}
 f(jk, j'k') &= n(jk, j'k') - \frac{n(jk)n(j'k')}{n} \\
 g(jk, j'k') &= \sum_{i=1}^k \frac{n(ijk)n(i'j'k')}{n(i)} - \frac{n(jk)n(j'k')}{n}
 \end{aligned}$$

이다.

그러므로 상관비 η^2 은

$$\eta^2 = \frac{SSB}{SST} = \frac{\sum_{j=1}^R \sum_{k=1}^{c_j} \sum_{j'=1}^R \sum_{k'=1}^{c_{j'}} f(jk, j'k') a_{jk} a_{j'k'}}{\sum_{j=1}^R \sum_{k=1}^{c_j} \sum_{j'=1}^R \sum_{k'=1}^{c_{j'}} g(jk, j'k') a_{jk} a_{j'k'}}$$

로 나타난다.

각 증후에 대하여 환자는 하나의 범주에 해당하므로

$$\sum_{k=1}^{c_j} \delta_{i\alpha}(jk) = 1$$

이 성립하여 상관비를 최대로 하는 $\{a_{jk}\}$ 의 값이 유일하게 정해지지 않으므로

$$a_{j1} = 0, j = 1, 2, \dots, R \quad (3.1)$$

을 가정하여 위의 부정성을 제거하고자 한다.

식(3.1)의 조건하에서

$$\frac{\partial \eta^2}{\partial a_{j'k'}} = 0$$

을 정리하면

$$\sum_{j=1}^R \sum_{k=2}^{c_j} \{g(jk, j'k') - \eta^2 f(jk, j'k')\} a_{jk} = 0, k' = 2, 3, \dots, c_{j'} \quad (3.2)$$

이 구해진다. 식(3.2)는 행렬을 이용하여 나타내면 다음과 같다.

$$(G - \eta^2 F) \mathbf{a} = 0 \quad (3.3)$$

이때

$$G = [g(jk, j'k')], F = [f(jk, j'k')], \mathbf{a} = [a_{jk}]$$

이다.

식(3.3)은 가변수 $\{\delta_{i\alpha}(jk)\}$ 를 이용한 경우의 정준판별분석과정에 해당하므로(Matthew,G.와 William,D., 1978), 우리가 구하고자 하는 $\{a_{jk}\}$ 는 최대고유치 η^2 에 대응하는 고유벡터이다.

이렇게 구한 $\{a_{jk}\}$ 를 이용하여 증후와 증형을 수량화할 수 있다. 즉, 증형의 각 범주에 부여하는 수량화값 y_i 와 i 군의 증형의 환자 α 의 증후 j 의 수량화값 $x_{i\alpha}(j)$ 는

$$y_i = \frac{1}{n_i} \sum_{j=1}^R \sum_{k=1}^{c_j} a_{jk} \delta_{i\alpha}(jk) \tag{3.4}$$

$$x_{i\alpha}(j) = \sum_{k=1}^{c_j} a_{jk} \delta_{i\alpha}(jk)$$

이다. 위의 방법에 의해 수량화된 자료를 이용하여 증후로부터 증형을 진단할 수 있는 진단모형을 구축 하게 된다. 이렇게 구축된 진단모형은 한의전문가 시스템의 기초가 될 것이다.

4. 예제: 증풍자료

아래 자료는 경산대학교 부속 대구한방병원에서 1996년 9월에서 11월까지 수집한 증풍 환자 45명의 5개 증형과 8개 증후에 대한 임상자료이다. 자료의 공변량의 내용은 (표 4.1)에 나타나있다.

(표 4.1) 증풍자료의 공변량

증	범주1	범주2	범주3	범주4	범주5
현운	약간있다	없다			
갈음	구간	구건	정상		
체형	건장비만	물살비만	보통	허약수척	견실수척
담성	심하다	약간있다	없다		
소변색	단적	보통	청장		
한출	도한	정상	자한		
설질	홍	정상	담백		
설태색	무	황태	백태		
증형	화열	음허양항	담화	습담	기혈구허

위의 증형과 증후들에 대하여 가변수를 주어 식(3.3)에 의한 정준계수를 구한 결과가 (표 4.2)이다. 자료분석의 출력은 통계프로그램 SAS에 의한 것이다.

(표 4.2) 증풍자료의 가변수에 대한 정준계수

증	제1정준계수	제2정준계수	제3정준계수	제4정준계수
증형1	1.11	1.67	1.92	1.81
증형2	-0.23	3.12	-0.38	1.71
증형3	3.05	1.80	-0.71	1.26
증형4	0.68	0.67	-0.11	3.13
현운1	0.75	0.16	0.50	-0.11
갈음1	0.48	0.79	-0.29	0.95
갈음2	0.37	0.46	0.17	0.07
체형1	0.68	0.54	1.94	3.40
체형2	0.18	0.28	0.23	3.97
체형3	0.16	0.79	0.61	3.62
체형4	-1.63	0.38	0.35	2.98
담성1	1.56	-0.89	-1.38	-2.33
담성2	0.93	-0.12	-1.87	0.22
소변색1	0.84	0.05	-0.95	0.08
소변색2	-0.10	-0.76	-1.43	0.96
한출1	-0.61	0.62	1.83	-0.62
한출2	-1.00	0.68	1.90	-0.37
설길	0.06	0.43	-0.16	0.22
설태색1	-1.01	2.01	0.19	-0.30
설태색2	0.26	0.56	0.17	0.53

증후와 증형에 대하여 (표4.2)와 식(3.4)를 이용하여 수량화를 실시한 결과를 정리하면 (표 4.3)이 된다.

(표4.3) 중풍자료의 수량화

증	범주	제 1축	제 2축	제 3축	제 4축
증형	화열	0.25	0.23	1.18	0.03
	음허양항	-0.90	1.41	-0.48	-0.02
	담화	1.92	0.34	-0.71	-0.26
	습담	0.12	-0.58	-0.28	0.73
	기혈구허	0.71	-1.12	-0.20	-0.93
현운	없다	-0.27	-0.06	-0.18	0.04
	약간있다	0.48	0.10	0.32	-0.07
갈음	정상	-0.15	-0.21	0.00	-0.14
	구건	0.22	0.25	0.17	-0.06
	구갈	0.33	0.58	-0.29	0.81
체형	건실수척	0.14	-0.55	-0.61	-3.46
	허약수척	-1.49	-0.17	-0.26	-0.48
	보통	0.30	0.24	0.00	0.16
	물살비만	0.32	-0.27	-0.38	0.51
	건장비만	0.82	-0.01	1.33	-0.06
담성	없다	-0.30	0.07	0.52	0.05
	약간있다	0.63	-0.05	-1.35	0.27
	심하다	1.26	-0.82	-0.86	-2.28
소변색	청장	-0.18	0.46	1.16	-0.60
	보통	-0.28	-0.30	-0.27	0.36
	단적	0.66	0.51	0.21	-0.52
한출	자한	0.90	-0.63	-1.77	0.37
	정상	-0.10	0.05	0.13	0.00
	도한	0.29	-0.01	0.06	0.25
설질	정상	-0.03	-0.19	0.07	-0.10
	홍	0.03	0.24	-0.09	0.12
설태색	백태	-0.04	-0.55	-0.12	-0.27
	황태	0.22	0.01	0.05	0.26
	무	-1.05	1.46	0.07	-0.57

참고문헌

1. 노형진(1990).다변량해석, 석정, 서울.
2. 서산영웅(1980).한방의학의 기초와 진료, 창원사, 동경.
3. 양사두, 진귀연(1991). 실용중서의 결합진단 치료학, 중국의약과기출판사, 북경.
4. 허명희(1992). 수량화 방법론의 이해, 자유아카데미, 서울.
5. Matthew, G. and William, D. (1978).*Discrete Discriminant Analysis*, John Wiley, New York.
6. Shin, Y. (1997). 한의진단모델의 추론과정에서 발생하는 불확실한 진단지식의 처리,*Journal of Statistical Theory & Methods*, Vol8, No1, 1-7.

A study on the quantification for Oriental Medicine Data

Yang-kyu Shin

Abstract

In oriental medicine, it is required that correct medical knowledge should be maintained for medical expert system which analyzes and diagnoses patients symptoms. Typical medical expert system has a knowledge base as its core, and the knowledge base contains a domain specific knowledge about patients records. However, oriental medicine diagnostic knowledge is formed mostly as qualitative data, knowledge could be ambiguous and uncertain. In this paper, we looked at quantification methods and propose a method for quantifying the oriental medicine diagnostic knowledge, which is improving the knowledge base of an oriental medicine expert system.

Key Word and Pharses: quantification method, qualitative data, oriental medicine diagnostic knowledge.