

FCM 알고리즘과 퍼지 소속도를 이용한 지능형 자가 진단 시스템

김광백
신라대학교 컴퓨터정보공학부
(gbkim@silla.ac.kr)

김주성
신라대학교 간호학과
(kimjusung@silla.ac.kr)

본 논문에서는 전문적인 지식이 부족한 일반인들을 대상으로 자신의 건강 상태를 파악 할 수 있는 지능형 자가 진단 시스템을 제안한다. 제안된 자가 진단 시스템은 보건 복지부에 제출된 '한국인이 부담을 가지는 질병' 관련 보고서를 참조하여 선정된 30가지의 질병과 각 질병에 대한 대표 증상을 이용하여 질병을 도출한다. 본 논문에서는 개선된 FCM 알고리즘을 적용하여 질병 종류를 군집화하고 각 질병의 증상과 관련된 질의 결과를 입력 벡터로 적용하여 사용자의 건강 상태를 진단한다. 기존의 방법에서는 입력 벡터와 군집 중심과의 거리를 측정 한 후 거리가 가까운 5가지를 선택하기 때문에 선택된 질의와 관련 없는 질병을 도출하는 단점이 있었다. 이러한 단점을 개선하기 위해, 선택된 질의와 도출된 질병에 대한 퍼지 소속도를 이용하여 정렬한다. 정렬된 질병에서 상위 5가지를 도출한 결과, 선택된 질의와 관련된 질병만을 도출하는 것을 확인 할 수 있었다.

논문접수일 : 2007년 01월

게재확정일 : 2007년 03월

교신저자 : 김광백

1. 서론

환자의 건강사정은 질병에 대한 이해와 치료방향을 설정하는 가장 기초적이며 필수적인 중요 단계이다. 그러나 의료자원에 대한 지리적인 거리나 심리적 부담감, 경제적 비용 등은 의료 수요자인 환자들의 의료자원 접근성에 많은 제약을 초래해 왔다.

의료시설이 도시지역에 편중됨에 따라 도서산간 지역주민은 의료자원의 선택범위에서 불평등을 경험하고 있으며 긴 진료 대기 시간과 짧은 진료 면담은 질병에 대한 충분한 정보 교류를 원하는 환자의 기대를 충족시키지 못하고 있다. 특히,

진료분야에 대한 환자들의 정보 부족은 잘못된 진료 영역의 선택과 그에 따른 시간적, 경제적 손실의 많은 불편을 유발함에 따라 우선적으로 개선되어야 할 의료서비스 영역이기도 하다.

이와 같은 의료자원에 대한 접근성의 제약은 환자들이 자신의 건강 관리에 참여하는 의욕을 떨어뜨리며 이로 인해 자신의 건강상태를 파악하기 어렵고, 자신의 질병에 소홀하거나 정기적인 검진을 받지 못하게 되어 쉽게 치료 가능한 질병 조차도 큰 질병으로 악화되는 경우가 발생하고 있다. 따라서 이러한 문제를 해결하기 위해 활용 가능성이 높은 의료서비스의 일환으로 자신의 건강 상태를 스스로 진단하고 관리 할 수 있게 유도하는 자가

진단 시스템의 개발이 요구된다.

의료 정보 관련 소프트웨어는 자가 진단 시스템을 광범위한 의료 현장에 쉽게 적용 시킬 수 있는 유용한 방법이다[1]. 환자는 의료진과의 직접 대면에 따른 부담감 없이 그래픽 또는 영상 처리 기법으로 구현된 인체 구조나 질의를 통하여 일련의 자기 보고식 응답 절차를 거치면서 건강상태에 대한 평가가 이루어진다.

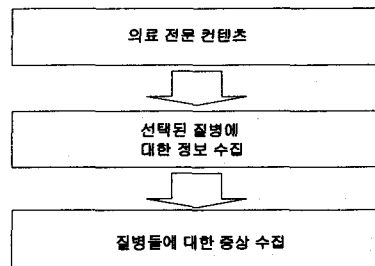
자가 진단 시스템은 정보의 통합성을 향상시킴에 따라 질의 응답간 면밀한 연계의 규명과 건강 문제의 우선순위를 결정하는데 도움을 준다. 또한 건강 정보 전달이 컴퓨터로 처리된 반응에 따라 개별화 되어 추후 의료진과의 심층 면접시 불필요한 인터뷰 과정 없이 의사 소통을 명확히 하여 빠르고 예측적인 건강 평가를 가능하게 한다[2]. 이는 시간적, 비용적 측면에서의 건강평가에 대한 효율성과 접근성을 높여 환자의 건강 증진뿐만 아니라 전문 영역별 의료진 간의 협력을 효과적으로 유도하여 장기적 측면에서는 의학적 치료의 장기화를 예방하고 의료비 지출을 줄이며 의료 서비스에 대한 만족도를 향상 시킬 수 있다[3]. 따라서 자가 진단 시스템을 통한 포괄적인 건강관리를 위해서는 수집된 자료에 대한 정보 처리 과정의 체계적인 개발과 그 타당성 검증이 시급한 과제라 할 수 있다.

이에 본 논문에서는 질병관리 부담이 큰 한국인의 호발 질병에 대한 지능형 자가 진단 시스템을 제안하고 개선된 FCM 알고리즘을 제시하여 각 질병들과 유사도를 측정하고, 측정된 유사도와 증상 따른 질병별 소속도를 이용하여 가장 큰 소속도와 유사성을 가지는 질병을 최종 결과로 도출 시킴으로써 제안된 지능형 자가 진단 시스템의 실용성을 검증한다.

2. FCM 알고리즘과 퍼지 소속도를 이용한 지능형 자가 진단 시스템

2.1 질병 및 증상 조사

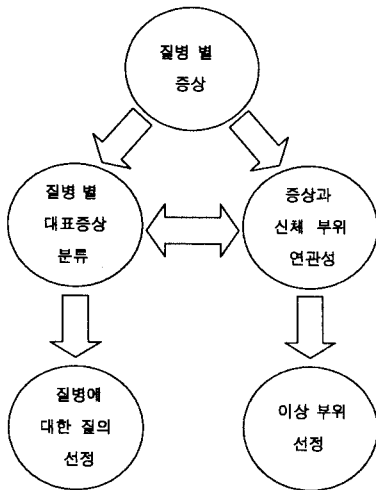
본 논문에서는 보건복지부에 제출된 ‘한국인의 질병 부담 2005년 보고서’를 참조하여 일반인에게 호발하며 질병관리가 부담되는 30종의 질병을 선정하였다. 질병 장애로 인한 부담이 가능 큰 경우는 당뇨병이었으며 다음으로 위·십이지장궤양, 천식, 뇌혈관질환, 류마티스성 관절염, 우울증 순으로 조사되었다[4]. 호발 위험 질병은 연령에 따라 차이가 있었는데 20대는 신경정신질환(우울증), 30~40대 위궤양과 간경화, 50~60대 암(신생물), 70대 이후는 뇌졸중, 심근경색증이 1위로 보고 되었으며[5] 이들 자료는 한국인의 10대 사망원인 질병과도 관련이 있었다. 한국인의 2005년 사망원인 질병으로 1위는 암(신생물)이었으며 순환기계질환(뇌혈관, 심혈관질환), 호흡기계질환(만성하기도질환, 폐렴), 내분비대사질환(당뇨병), 소화기계질환(간질환)의 순으로 나타났다[6].



[그림 1] 질병 증상 수집 과정

본 논문에서는 지능형 자가 진단 시스템을 제안하기 위해 각 질병에 대한 증상들은 의료 전문 콘텐츠 20곳에서 참조하여 수집한 후, 이를 바탕으로 각 질병에 대한 증상을 질의로 제시하였다[7, 8].

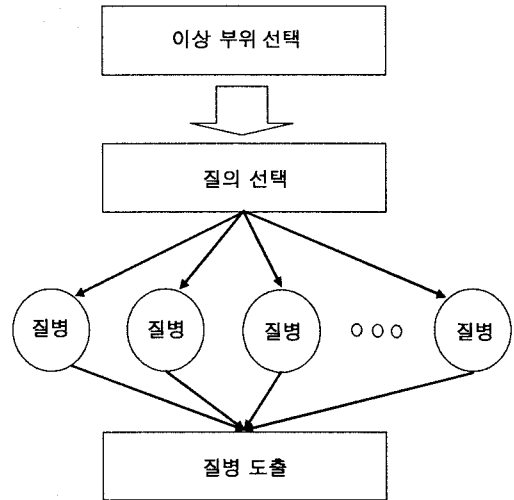
수집 과정은 [그림 1]과 같다. 질병 증상에 대한 질의는 가장 뚜렷하게 나타나는 증상들을 대표 증상으로 분류하고 이상 부위는 증상과 신체 부위의 연관성을 고려하여 채택한다. 본 논문에서 제안한 이상 부위와 질의 선정 구조는 [그림 2]와 같다.



[그림 2] 이상 부위와 질의 선정

2.2 제안된 지능형 자가 진단 시스템

제안된 지능형 자가 진단 시스템에서 사용자에게 진단 결과를 제공하는 과정은 다음과 같다. 13가지로 설정한 이상 부위가 표시된 인체도를 사용자에게 제공하여 자신이 아픈 부위를 선택하게 한다. 이상 부위를 선택하면 해당 부위에 발생 할 수 있는 증상들을 질의로 제공하고 질의에 대한 결과들을 패턴화 한다. 즉 선택된 질의 결과에 따라 각 질병들에 대한 소속 함수를 설계하여 소속도를 계산한 후, 계산된 소속도로 질의와 관련된 질병만을 선택하고 개선된 FCM 알고리즘을 적용하여 유사성을 측정 한 후, 최종 질병을 도출한다. 제안된 지능형 자가 진단 시스템의 구성은 [그림 3]과 같다.

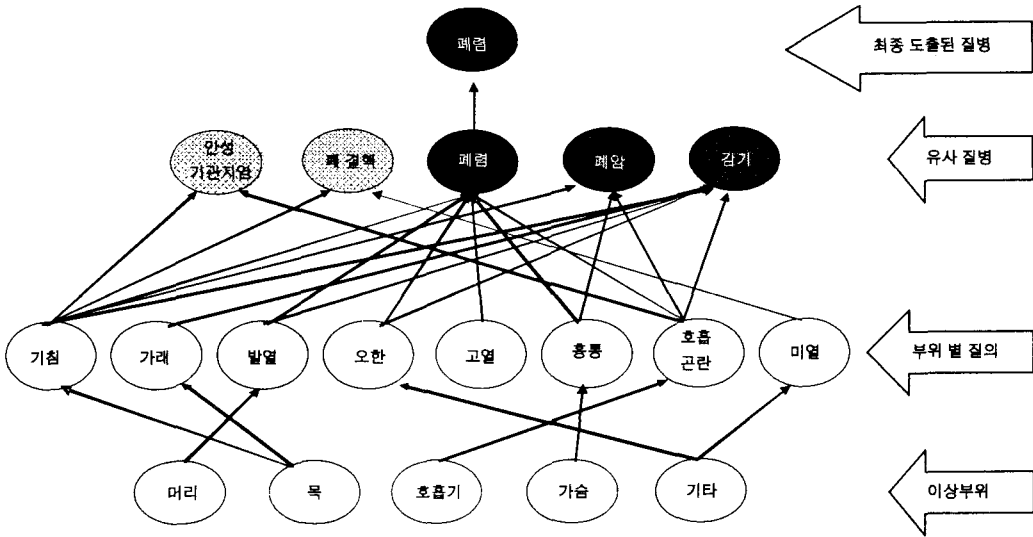


[그림 3] 제안된 지능형 자가 진단 시스템

질의 선택에 따른 질병 도출은 총 30종의 질병에 대해서 각 질병들의 증상에 대한 질의 결과를 입력 벡터로 제시 한 후, 소속 함수와 개선된 FCM 알고리즘을 적용하여 질병을 도출한다. 질병 도출 과정은 [그림 4]와 같다. [그림 4]에서 유사한 질병을 나타내는 층은 음영이 어두운 색상 일수록 입력되는 질의들 중 많은 질의를 포함하는 군집을 의미한다. 본 논문에서 제시한 개선된 FCM 알고리즘은 군집들의 대칭성 측도에 퍼지 이론을 적용하여 기존의 FCM 알고리즘으로 군집화 한 결과를 재 군집화 함으로서 군집화의 정확성을 개선한다.

2.3 개선된 FCM 알고리즘

제안된 자가 진단 시스템은 사용자의 증상과 질병의 유사성을 측정하기 위하여 개선된 FCM 알고리즘을 적용한다. 기존의 FCM 알고리즘은 입력 벡터들과 각 군집 중심과의 거리를 이용하여 측정된 유사도에 기초한 목적 함수의 최적화 방식을 사용한다[9]. 그러나 측정된 패턴과 군집 중심과의



[그림 4] 최종 질병 도출 과정

거리만을 이용하여 군집화하는 알고리즘이기 때문에 군집 공간상의 패턴들의 분포에 따라 바람직하지 못한 군집화 결과를 보일 수 있다[10, 11]. 예

과 같다.

식 (1)에서 $\text{deg}(x_i, x_j, c)$ 는 점 c 를 중심으로 점 x_i 와 x_j 의 각도를 의미하며 $\text{ratio}_d(x_i, x_j, c)$ 는 식(?)

$$\text{Symmetric}(x_i, c) = \max_{j \in \forall \text{pattern}, i \neq j} \left(\begin{array}{l} (1-\alpha) \left(1 - \frac{\text{deg}(x_i, x_j, c)}{180} \right) \\ - (\alpha * \text{ratio}_d(x_i, x_j, c)) \end{array} \right) \quad (1)$$

를 들면 군집 공간상에 분포된 패턴들이 타원의 형태를 가지는 경우이거나 패턴의 분포가 군집 경계면에 존재하고 그 기본형이 서로 교차하는 경우를 들 수 있다. 이러한 경우에는 기본형의 중심로부터 패턴간의 거리를 측정하는 유클리드 거리법 적용 시 올바르지 못한 군집화 결과의 원인이 될 수 있다. 따라서 본 논문에서는 대칭적 성질을 이용하는 대칭성 측도에 퍼지 이론을 적용하여 군집간의 거리에 따른 변화와 군집 중심의 위치, 그리고 군집 형태에 따라 영향을 덜 받는 개선된 FCM 알고리즘의 군집화 방법을 적용한다. 개선된 FCM 알고리즘에서 적용하는 대칭성 측도는 식 (1)

와 같고, α 는 퍼지 이론을 이용한 가중치이다. 가중치 계산은 식 (3)과 같다.

$$\text{ratio}_d(x_i, x_j, c) = \left\{ \begin{array}{l} \frac{d(x_j, c)}{d(x_i, c)} \quad \text{if}(d_i > d_j) \\ \frac{d(x_i, c)}{d(x_j, c)} \quad \text{if}(d_i < d_j) \end{array} \right\} \quad (2)$$

여기서 $d(x_j, c)$ 는 유클리드 거리를 의미한다.

$$\alpha = \frac{d(c_i, c_j)}{\sqrt{D_m}} \quad (3)$$

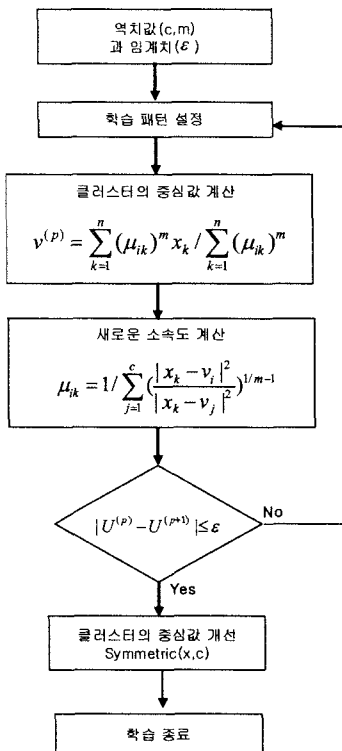
식 (1)을 이용하여 대칭성 측도를 구한 상수값을 $\mu(x)$ 라고 정의한 후, 군집의 중심 값을 계산한다. 군집의 중심 값 계산은 식 (4)와 같다.

$$v^{(p)} = \frac{\sum \mu(x)x}{\sum \mu(x)} \quad (4)$$

개선된 FCM 알고리즘으로 군집화 된 각 군집들의 중심 값과 현재 패턴 사이의 유사도 U 를 식 (5)를 적용하여 계산한다.

$$U = \sqrt{\sum_{i=0}^{k-1} (x_i - c_i)^2} \quad (5)$$

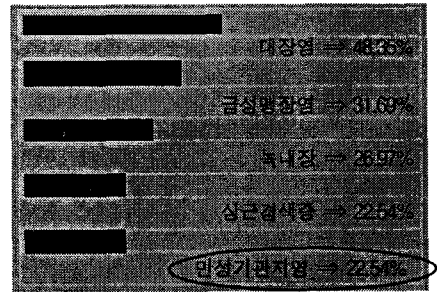
개선된 FCM 알고리즘의 학습 과정은 [그림 5]와 같다.



[그림 5] 개선된 FCM 알고리즘

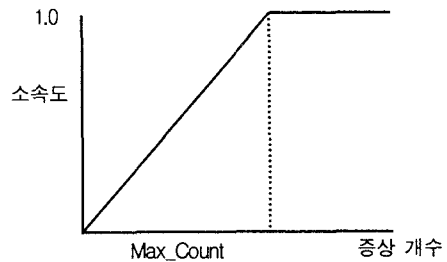
2.4 증상에 따른 질병 별 소속도

제안된 자가 진단 시스템에 개선된 FCM 알고리즘을 적용한 결과, [그림 6]과 같이 연관성이 적은 질병을 결과로 도출하는 단점이 있다.



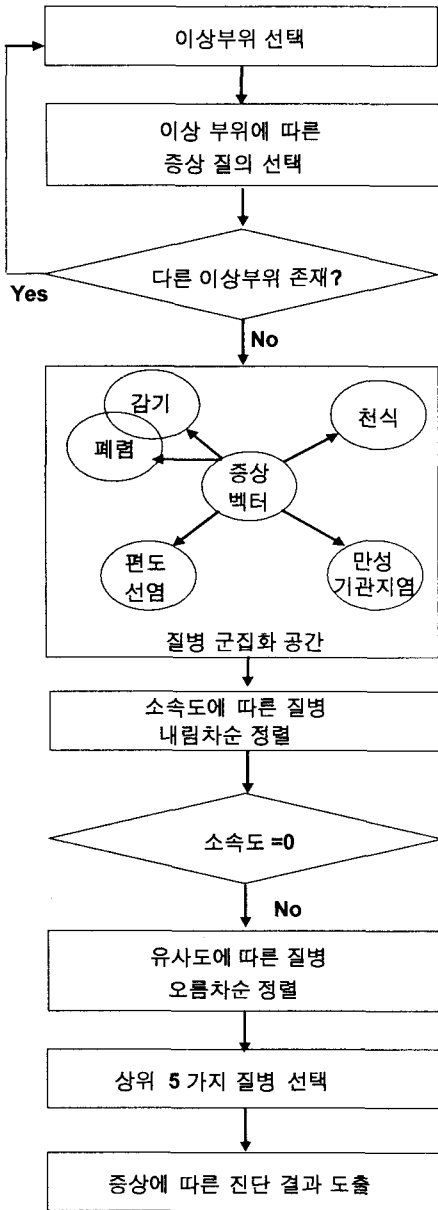
[그림 6] 개선된 FCM 알고리즘을 이용한 질병 도출 화면

따라서 자가 진단 방법의 단점을 개선하기 위해, 본 논문에서는 증상에 따른 질병 별 소속도를 계산한다. 각 질병들이 가질 수 있는 증상과 선택한 질의와 일치하는 증상의 수를 계산하고, 각 질병이 가질 수 있는 증상의 최대 개수를 계산한 후, 제안한 소속 함수에 적용하여 선택된 질의에 따른 질병별 소속도를 계산한다. 증상에 따른 질병별 소속함수는 [그림 7]과 같다.



[그림 7] 증상에 따른 질병 별 소속함수

여기서 Max_Count는 각 질병이 가질 수 있는



[그림 8] 제안된 지능형 자가 진단 시스템의 전체적인 구성도

증상의 총 개수 이다. 질병마다 가지는 증상이 다르기 때문에 Max_Count 값은 질병마다 달라진다.

그리고 소속함수의 입력 값은 선택한 질의와 질병의 증상이 일치하는 개수이다. 소속 함수에서 소속도 계산은 식 (6)과 같다.

If (입력 값 < 0) then $\mu = 0$
 Else If (입력 값 > 0 AND 입력 값 < Max_Count) then

$$\mu = \frac{Max_Count - \text{입력 값}}{Max_Count}$$

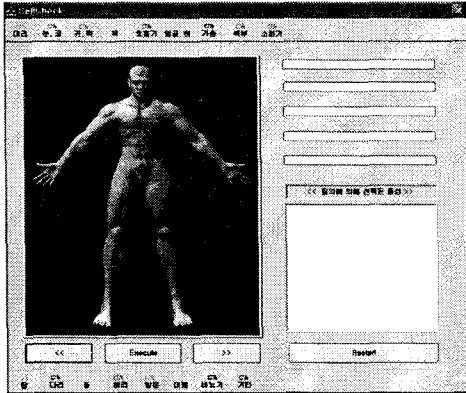
 Else If (입력 값 > Max_Count) then $\mu = 1$ (6)

제안된 자가 진단 시스템에서는 소속도를 이용하여 질병들을 내림 차순 정렬 한 후, 소속도가 0이 아닌 질병들을 개선된 FCM 알고리즘에 적용하여 유사도 U를 측정 한 후, 총 5종의 질병을 최종 도출한다. 최종 5종의 질병 도출 과정은 [그림 8]과 같다

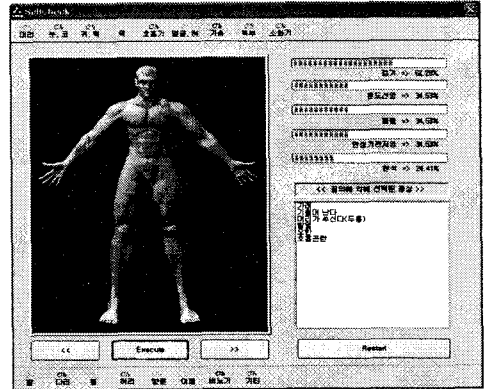
3. 실험 및 결과 분석

실험 환경은 Intel Pentium_IV 2GHz CPU와 256MB RAM이 장착된 IBM 호환 PC상에서 VC++ 6.0으로 실험 하였다. 제안된 자가 진단 시스템은 [그림 6]과 같은 구조로 구현하였다. 실험에 사용된 질병과 질의는 각각 30종과 125종이고 [그림 9]는 제안된 자가 진단 시스템의 초기 화면이다.

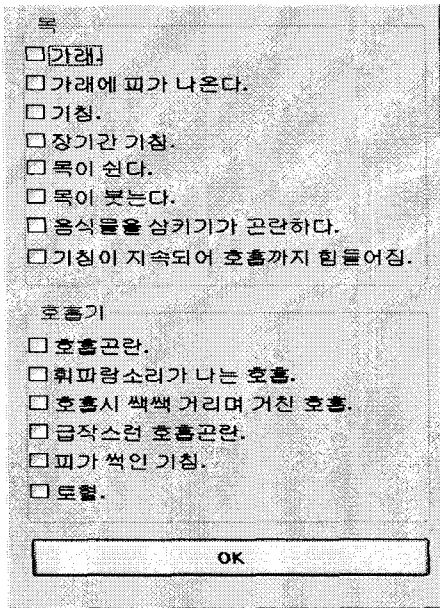
[그림 9]의 인체도에 설정된 각 부위를 선택하거나 위, 아래에 나타나 있는 부위 별 버튼을 선택하면 해당 부위들에 대한 질의가 나타난다. [그림 10]은 호흡기와 목에서 발생할 수 있는 증상에 대한 질의 화면이다.



[그림 9] 제안된 지능형 자가 진단 시스템의 초기 화면



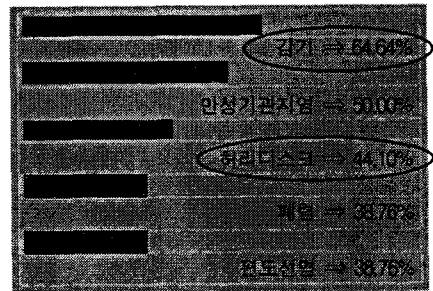
[그림 11] 최종 결과 도출 화면



[그림 10] 호흡기와 목에 대한 질의 화면

부위별로 자신의 증상에 해당하는 질의를 선택한다. 질의가 선택되면 그 결과를 특정 벡터로 제시하여 최종적으로 사용자가 가질 수 있는 질병 5종을 도출하게 된다. 최종 결과 도출 화면은 [그림 11]과 같다.

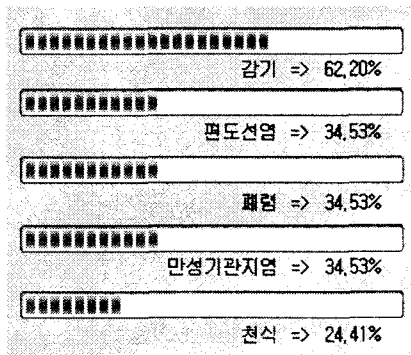
제안된 질병 자가 진단 시스템은 증상에 따른 질병을 도출하기 위해 개선된 FCM 알고리즘과 증상에 따른 질병 별 소속도를 적용하였다. [그림 12]와 [그림 13]은 같은 증상에 대해 각각 FCM 알고리즘과 개선된 FCM 알고리즘을 적용한 결과 화면이다. 실험에 적용한 증상은 가래, 기침, 두통, 발열, 오한, 호흡 곤란이다.



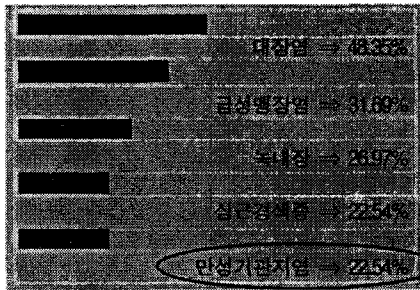
[그림 12] FCM 알고리즘을 적용한 자가 진단 결과 화면

[그림 12]에서는 증상들과 연관성이 적은 질병인 ‘허리디스크’가 3번째로 높은 유사도를 나타내지만, [그림 13]에서와 같이 개선된 FCM 알고리즘을 적용하였을 경우에는 같은 증상에 대해 연관성이 큰 질병들만이 도출되었고, 증상과 질병의 유사

도 또한 개선되었음을 확인 할 수 있었다. 그러나 개선된 FCM 알고리즘만을 적용한 자가 진단 시스템에서 만성 설사, 탈수 상태, 구토, 복통 증상에 대한 결과를 도출한 경우에는 [그림 14]와 같이 연관성이 적은 ‘만성기관지염’이 5번째로 높은 유사도를 나타내었다.



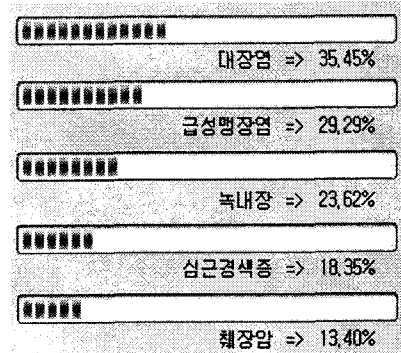
[그림 13] 개선된 FCM 알고리즘을 적용한 자가 진단 결과 화면



[그림 14] 개선된 FCM 알고리즘을 적용하여 연관성이 적은 결과 도출 화면

개선된 FCM 알고리즘을 적용한 자가 진단 시스템의 문제점을 개선하기 위하여 증상에 따른 질병 별 소속도를 적용한 결과는 [그림 15]와 같다.

[그림 14]와 [그림 15]를 분석한 결과, 소속도를 적용하여 5번째로 높은 유사도를 나타내는 질병이 증상과 연관성이 있는 질병을 도출하여 개선된



[그림 15] 증상에 따른 질병 별 소속도를 적용한 결과 화면

FCM 알고리즘만을 적용한 자가 진단 시스템의 결과보다 개선되는 것을 확인 할 수 있다. 제안한 지능형 자가 진단 시스템의 질병 군집화에 적용된 개선된 FCM 알고리즘의 초기 파라미터 설정은 <표 1>과 같다.

<표 1> 개선된 FCM 알고리즘 초기 설정 파라미터

	패턴수	입력 노드수	m	군집 개수
개선된 FCM 알고리즘	50	125	1000	30

<표 1>에서 m은 개선된 FCM 알고리즘의 지수 가중치로 패턴을 표현하기 위한 공간의 차수이다. 패턴은 각 질병이 가지는 증상들의 정보로 구성하였으며, 입력 노드는 사용자가 선택한 증상을 패턴화 한 것이다. 제안된 지능형 자가 진단 시스템에서 군집은 질병이 가지는 모든 증상을 의미하며 그 개수는 30개로서 총 질병 수와 같다.

4. 결론

본 논문에서는 일상생활에서 건강 관리의 여유

가 부족하거나 도심지역에 비해 상대적으로 의료 혜택의 기회가 적고 의료 시설이 부족한 도서 산간 지방 주민, 그리고 전문적인 지식이 부족한 일반인들을 대상으로 자신의 건강 상태를 파악 할 수 있는 지능형 자가 진단 시스템을 제안하였다.

본 논문에서는 질병관리 부담이 큰 한국인의 호흡기 질환에 대한 지능형 자가 진단 시스템을 제안하고 개선된 FCM 알고리즘을 제시하여 각 질병들과 유사도를 측정하고, 측정된 유사도와 증상에 따른 질병 별 소속도를 이용하여 가장 큰 소속도와 유사성을 가지는 질병을 최종 결과로 도출하였다.

기존의 웹 상에서 제공하는 자가 진단 방법은 사용자의 증상에서 유추 할 수 있는 모든 질병들을 결과로 제공하는 것으로서 신뢰성과 효율성이 결여되었다. 이에 반해 개선된 FCM 알고리즘만을 적용한 자가 진단 시스템은 사용자가 가질 수 있는 질병들에 대한 소속 정도를 제공함으로써 신뢰성과 효율성을 가지며 또한 사용자로 하여금 자신이 가질 수 있는 질병에 좀더 신중을 기할 수 있게 유도하는 장점을 가진다. 그러나 개선된 FCM 알고리즘만을 적용할 경우에는 증상에 대한 질병 결과를 도출하는 과정에서 증상과 질병간의 유사도를 단지 유클리드 거리만을 적용하여 측정 하기 때문에 증상과 연관성이 적은 질병도 결과 중의 하나로 도출되는 경우가 발생하였다. 따라서 이러한 문제점을 개선하기 위해 본 논문에서는 증상에 따른 질병별 소속도를 적용하여 증상과 연관성 있는 질병만 도출할 수 있도록 개선하였다.

참고문헌

- [1] Mullen K. H. Berry D. L., and Zierler B. K., "Computerized symptom and quality of life assessment for patients with cancer part II; acceptability and suability", *Oncology nursing forum*, Vol.31, No.5(2004), E84-89.
- [2] Berry D. L., Trigg L. J., Lober W. B., Karras B. T., Galligan M. L., Austin-Seymour M., and Martin S., "Computerized symptom and quality of life assessment for patients with cancer part I; development and pilot testing", *Oncology nursing forum*, Vol.31, No.5(2004), E75-83.
- [3] Ford D. F., Pincus H.A., Unützer J, Bauer M. S., Gonzalez J. J., and Wells K. B., "Practice based interventions", *Mental Health Services Research*, Vol.4, No.4(2002), 99~204.
- [4] 통계청, 사망원인 별 질병(2005년 12월), Retrieved Oct 1, 2006 from http://kosis.nso.go.kr/cgi-bin/sws_999.cgi
- [5] <http://www.mohw.go.kr>
- [6] WHO.ICD-10 International Statistical Classification of Diseases and Related Health Problems, tenth revision. Vol.2(1993), Geneva, WHO.
- [7] Simans, David G., Travell, Janet G. and Simans, Lois S., *Myofascial Pain and Dysfunction, The Trigger Point Manual*, Vol.1, Vol.2(1983), Williams & Wilkins.
- [8] 정태준, *질병 자가진단*, 장원출판사, 1995.
- [9] Arun D. K., *Computer Vision and Fuzzy Neural Systems*, Prentice Hall PTR, 2001.
- [10] J. Bezdek, "A convergence theorem for the fuzzy ISODATA clustering algorithm", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol.PAMI2, No.1(1980), 1~8.
- [11] Kim K. B., Lee D. U., Sim K. B., "Performance Improvement of Fuzzy RBF Networks", *Lecture Notes in Computer Science*, LNCS 3610, Springer, 237~244, 2005.

Abstract

An Intelligent Self Health Diagnosis System using FCM Algorithm and Fuzzy Membership Degree

Kwang-Baek Kim* · Ju-Sung Kim**

This paper shows an intelligent disease diagnosis system for public. Our system deals with 30 diseases and their typical symptoms selected based on the report from Ministry of Health and Welfare, Korea. Technically, the system uses a modified FCM algorithm for clustering diseases and the input vector consists of the result of user-selected questionnaires. The modified FCM algorithm improves the quality of clusters by applying symmetrically measure based on the fuzzy theory so that the clusters are relatively sensitive to the shape of the pattern distribution. Furthermore, we extract the highest 5 diseases only related to the user-selected questionnaires based on the fuzzy membership function between questionnaires and diseases in order to avoid diagnosing unrelated disease.

Key words : Intelligent Disease Diagnosis System, Modified FCM Algorithm, Pattern Distribution, Fuzzy Membership Function

* Division of Computer and Information Engineering, Silla University

** Dept. of Nursing, College of Medical Life Science, Silla University