

Diagnosis of Pet by Using FCM Clustering

Kwang-Baek Kim*

*Professor, Dept. of Artificial Intelligence, Silla University, Busan, Korea

[Abstract]

In this paper, we propose a method of disease diagnosis system that can diagnose the health status of household pets for the people who lack veterinary knowledge. The proposed diagnosis system holds 50 different kinds of diseases with the symptoms for each of them as a database to provide results from symptom input. Each disease database has its own symptom codes for a disease, and by using the disease database, FCM clustering technique is applied to disease which outputs membership degree to determine diseases close to the input symptom as a pet diagnosis result.

The implementation results of the proposed pet diagnosis system were obtained by the number of selected symptoms and the possibility values of the diseases that have the selected symptoms being sorted in descending order to derive top 3 diseases closest to the pet's symptom.

▶ **Key words:** Diseases, Symptom, FCM Clustering, Membership Degree, Pet Diagnosis

[요 약]

본 논문에서는 가정에서 많이 기르는 반려견을 바탕으로 반려견 질병에 대한 전문적인 수의학 지식이 부족한 일반인들을 대상으로 자신의 반려견의 건강 상태를 파악할 수 있는 진단 시스템을 제안한다. 제안된 진단 시스템은 50가지 질병과 각 질병의 증상을 데이터베이스에 구축하여 입력된 증상을 통해서 반려견의 질병을 도출한다. 각 질병 데이터베이스에는 질병에 해당하는 증상 코드들을 가지고 있으며, 이러한 질병에 대한 데이터베이스를 이용하여 군집화 기법인 FCM 클러스터링 기법을 적용하여 질병을 클러스터링하고 그 결과 값인 소속도를 바탕으로 입력된 증상과 가까운 질병들을 도출하여 반려견의 진단 결과를 제공한다. 제안된 반려견 진단의 구현 결과에서는 선택한 증상들의 개수와 선택된 증상들이 포함된 질병들의 가능성 값을 구하여 내림차순으로 정렬하여 반려견의 증상과 가장 가까운 질병 상위 3가지를 도출하였다.

▶ **주제어:** 질병, 증상, FCM 클러스터링, 소속도, 반려견 진단

• First Author: Kwang-Baek Kim, Corresponding Author: Kwang-Baek Kim
*Kwang-Baek Kim (gbkim@silla.ac.kr), Dept. of Artificial Intelligence, Silla University
• Received: 2021. 01. 25, Revised: 2021. 02. 19, Accepted: 2021. 02. 23.

I. Introduction

사회가 고도로 발달되면서 물질이 풍요로워지는 반면, 인간은 점차 자기중심적이고, 마음은 고갈되어 간다. 이에 비해 동물의 세계는 항상 천성 그대로이며 순수하다. 사람은 이런 동물과 접함으로써 상실되어가는 인간 본연의 성정을 되찾으려 한다. 이것이 동물을 반려하는 일이며, 그 대상이 되는 동물을 반려동물이라고 한다. 동물이 인간에게 주는 여러 혜택을 존중하며 동물은 사람의 장난감이 아닌 더불어 살아가는 존재임을 의미한다[1,2]. 이렇듯 이제는 집에서 기르는 동물을 애완동물이라는 개념을 넘어서 반려동물이라고 한다. 즉, 집에서 기르는 가축이 아니라 집에서 같이 먹고 자며 살아가는 하나의 가족 구성원인 것이다. 이러한 반려동물이 질병에 걸려 아프게 되면 가족 구성으로서 상심이 크다. 사람이든 동물이든 질병을 초기에 발견하여 제때 조치를 취하면 그 만큼 수명을 연장시킬 수 있고, 질병을 완치할 가능성이 높아진다. 사람의 경우에는 상호간의 의사소통이 가능하여 질병을 초기에 발견할 수 있지만, 의사소통이 원활하지 못한 반려동물 같은 경우에는 반려동물의 행동이나 몸의 상태 변화를 확인하여 유추해야한다. 그러나 전문적인 수의학 지식이 부족한 일반인들은 어떠한 반려동물의 행동이나 몸의 상태 변화를 확인하여도 자신의 반려동물이 질병에 걸렸다는 것을 인지하지 못하거나 인지하더라도 어떻게 대처해야 할지 모른다. 이러한 행동과 몸의 변화가 나타내는 의미를 모르거나 자신의 반려동물이 보이고 있는 질병의 심각성을 몰라, 대수롭지 않게 생각하여 병원에 데리고 가지 않을 경우, 간단하게 치료 가능한 질병임에도 불구하고 시기를 놓쳐 반려동물이 큰 고통을 겪게 되고 심각하면 생명이 위험한 상황이 발생할 수도 있다[3]. 이러한 문제점을 개선하기 위해 본 논문에서는 가정에서 많이 기르는 반려견을 바탕으로 반려견 질병에 대한 전문적인 수의학 지식이 부족한 일반인들을 대상으로 반려견의 건강 상태를 파악할 수 있는 진단 시스템을 제안한다.

II. The Proposed Scheme

1. Diseases and symptoms

제안된 반려견 진단 시스템은 ‘(愛犬)질병의 지식과 길들이는 법 - 유원출판사’와 ‘애완견의 질병과 치료 - 하서출판사’에서 반려견 질병에 관련된 서적을 이용하여 질병을 수집하고 반려견 질병 중 50가지 질병들을 선정하여 선

정된 질병과 관련된 증상들을 내과, 몸, 얼굴, 다리와 같이 4가지 부위별로 구분하여 92개의 증상들을 선정하였다[4].

2. Database implementation

본 논문에서 설계한 질병, 증상의 테이블 구조는 표 1 및 표 2와 같다.

Table 1. Disease Table

ID	Disease	Symptom	Symptom codes
1	ascariasis	...	8-12 ...
2	ancylostomiasis	...	6-8 ...
.
.
50	diabetes mellitus	...	7-8-43

Table 2. Symptom Table

ID	Symptom	Codes of Pet body parts
1	Excessive sebum secretion	2
2	Cough	1
.
.
92	Ass drag	2

표 1은 질병 테이블로 각 ID, 질병명, 증상, 증상코드, 질병 설명으로 구성하였고, 질병 테이블의 증상코드는 표 2의 증상 테이블의 ID를 가리키는 것으로 해당 질병에 나타나는 증상들을 의미한다. 표 2는 증상 테이블로 ID, 증상명, 부위 코드로 구성하였다.

3. FCM Clustering

FCM은 하나의 클러스터에 속해져 있는 각각의 데이터 점을 소속 정도에 의해서 클러스터에 대한 데이터의 소속 정도를 일일이 열거한 데이터 분류 알고리즘이다[5]. 일반적으로 FCM 클러스터링 기법은 그림 1과 같은 과정으로 수행한다.

그림 1과 같이 FCM 알고리즘은 클러스터링 기법의 하나로 초기 클러스터링 알고리즘인 HCM(Hard-C Means) 알고리즘을 개선한 알고리즘이다[6,7]. 두 알고리즘의 가장 큰 차이는 FCM 알고리즘이 퍼지분할 기법을 적용한다는 점이다. 퍼지분할이란 0과 1사이의 소속정도에 의해서 주어진 데이터를 특정 다수의 그룹에 속하게 하는 기법이다. 따라서 FCM 알고리즘은 하나의 클러스터에 속해져 있는 각각의 데이터의 소속정도를 구하여 소속도를 기반으로 데이터들을 각각의 클러스터에 분류하는 알고리즘이다. FCM 알고리즘은 다음과 같다.

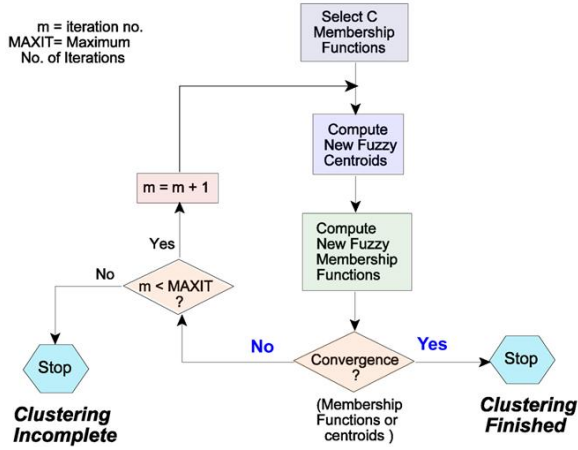


Fig. 1. FCM Clustering Processing Diagram

Step 1에서는 클러스터의 개수, $c(2 \leq c < n)$, 지수가 중치 $m(1 \leq m < \infty)$, 소속함수 $u^{(0)}$ 를 초기화한다.

Step 2에서 중심 벡터 v_{ij} 를 식 (1)과 같이 계산한다.

$$v_{ij} = \frac{\sum_{k=1}^n (u_{ik})^m x_{kj}}{\sum_{k=1}^n (u_{ik})^m} \quad \{v_i | i = 1, 2, \dots, c\} \quad (1)$$

Step 3에서는 새로운 소속 함수 $u^{(r+1)}$ 를 계산하기 위해 식 (2)와 같이 d_{ik} 를 계산하고 새로운 소속 함수 $u^{(r+1)}$ 를 계산한다. 여기서 c 를 클러스터 개수, d_{ik} 를 k 번째 패턴 x_k 와 i 번째 클러스터 중심 벡터 v_i 사이의 거리, u_{ik} 는 i 번째 클러스터에 속해 있는 패턴 중에 k 번째 패턴 x_k 의 소속도이다.

$$d_{ik} = d(x_k - v_i) = \left[\sum_{j=1}^l (x_{kj} - v_{ij})^2 \right]^{1/2} \quad (2)$$

식 (2)에서 l 은 입력 노드 수이고, 식 (3)에서 c 는 클러스터 개수이다.

$$u^{(r+1)*} = \frac{1}{\sum_{j=1}^c \left(\frac{d_{jk}^r}{d_{ik}^r} \right)^{2/m-1}} \text{ for } I_k = \emptyset \quad (3)$$

Step 4에서는 식 (4)와 같이 새로운 소속 함수(U^{r+1})와 이전 소속 함수(U^r)간의 변화율을 계산한다. 그리고 초기 클러스터 개수 설정에 따른 변화율을 누적한다. 변화율이 임계치(ε) 보다 차이 값이 크면 Step2로 차이 값이 적으면 Step 5로 간다.

$$\Delta = \| U^{(r+1)} - U^{(r)} \| = \max_{i,k} \left| u_{ik}^{(r+1)} - u_{ik}^{(r)} \right| \quad (4)$$

$$\text{if}(k > \varepsilon_t) k = \Delta + k$$

Step 5에서는 FCM 알고리즘을 적용하여 식 (5)와 같이 초기 클러스터의 개수를 설정한다.

$$\text{if}(c > 55) \quad N_c = \sum_{j=0}^h \Delta N_j, \quad N_e = \sum_{c=10}^{20} N_c \quad (5)$$

식 (5)에서 c 는 초기 클러스터의 개수이고 j 는 초기 클러스터의 수를 설정하여 클러스터링하여 생성되는 클러스터의 수이다. ΔN_j 는 생성된 클러스터들의 누적 변화율이고 N_c 는 초기 클러스터의 개수에 따른 클러스터의 변화율이다. N_e 는 설정된 초기 클러스터의 개수에 따른 클러스터들의 변화율을 모두 더한 값이다.

일반 증상들은 서로 다른 질병들의 일부 증상으로 나타나기 때문에 FCM에서 초기 클러스터의 개수를 각각 질병의 수에 +1에서 +5까지를 더한 범위에서 클러스터의 수를 설정하여 실험한 후에, N_e 값이 가장 적은 초기 클러스터를 가진 개수를 최종 클러스터로 설정하였다.

제안된 반려견 진단에 적용된 FCM 알고리즘은 선택한 증상과 클러스터의 증상이 일치하는 수를 FCM 알고리즘의 출력 값에 반영한다.

III. The Implementation and Result Analysis

본 논문에서 제안한 방법을 Intel(R) Core(TM) i5-8400 CPU, 8.00GB RAM이 장착된 PC상에서 jdk 12.0.2, Android Studio으로 구현하였다. 구현에 사용된 질병들은 전체 반려견 질병들 중에서 서로 다른 증상을 보이는 50개의 질병을 선정하였고 그 질병에 나타나는 증상 92개를 내부, 몸, 얼굴, 다리와 같이 4가지 부위별로 구분하였다. 제안된 반려견 진단 시스템의 초기 화면은 그림 2와 같다.

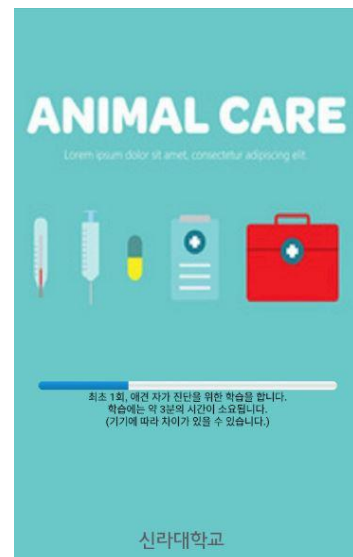


Fig. 2. Initial Screen of the Pet's Diagnosis

그림 3은 초기 화면 다음 화면으로 사용자가 증상을 선택한 화면이다. 선택 부위를 선택하면 해당 부위에 관련된 증상의 항목들이 나타나고 사용자가 증상을 선택한다. 그림 3에서 선택된 증상들은 ‘기침’, ‘흰자위 총혈’, ‘입냄새’, ‘체중 감소’, ‘기력 상실’, ‘끼억끼억하는 소리’이고 프로그램 최초 실행 시 클러스터링 하였던 소속도를 바탕으로 선택된 증상들의 소속도가 높은 질병이 해당 증상과 관련된 질병으로 판단하고, 최종 질병들을 도출한다.



Fig. 3. Selecting Body Part of the Pet

그림 4는 그림 3에서 선택된 증상들을 바탕으로 FCM 알고리즘을 적용하여 도출한 최종 결과이다.

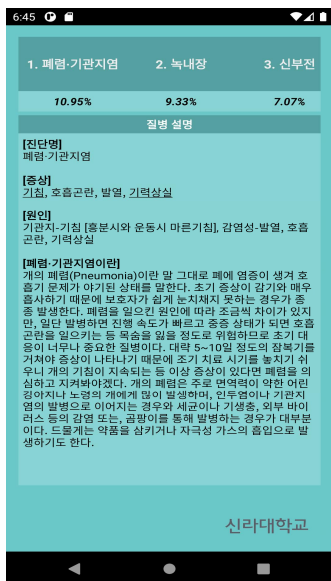


Fig. 4. Diagnosis result in Fig. 3

그림 4의 결과를 분석하면 표본 질병들이 증상을 기준으로 클러스터링 되었고 증상들은 소수의 질병에게만 속하는 것이 아니라 다수의 질병에 속한다. 따라서 본 논문에서는 FCM 알고리즘을 적용하여 클러스터의 소속도 값이 높은 값을 기준으로 질병들을 도출하였다. 따라서 그림 3과 같이 선택된 증상이 6 개인 경우에 대해 선택된 증상들이 특정 클러스터들에 포함되었기 때문에 소속도가 높은 질병에 해당되는 클러스터가 선택되었다. 선택하는 증상들이 질병과의 소속도가 높게 나타나는 경우에는 진단 결과에 대한 정확성이 높게 나타났다.

그림 5는 ‘기침’, ‘발육 부진’, ‘식욕항진’과 같은 증상을 선택한 화면이다.

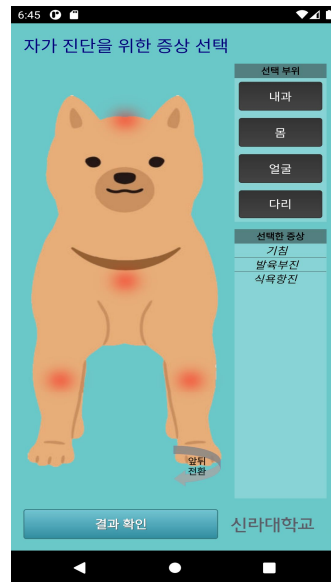


Fig. 5. An Example with a Small Number of Symptoms

해당 질병에 대한 증상의 개수가 적게 선택되면 FCM에서는 선택된 증상이 다른 질병에도 속하는 경우가 발생하고 선택된 증상이 다른 질병에서 소속도가 높게 나타나는 경우도 발생하므로 진단 결과의 정확성이 낮아진다. 따라서 그림 6과 같이 부정확하게 반려견의 진단 결과가 도출되는 경우는 해당 질병의 증상이 적게 선택되거나 대표 증상이 선택되지 않아서 클러스터의 소속도 값이 낮게 나타난 것으로 확인되었다.

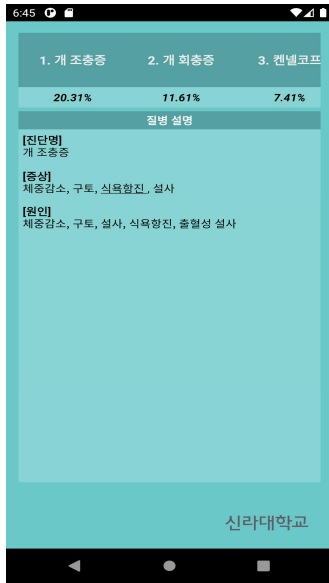


Fig. 6. Incorrect Diagnosis Result

녹내장 질병에 대해 사용자가 ‘흰자위 충혈’, ‘각막 혼탁’, ‘눈물 넘쳐흐름’을 선택하여 FCM 기법으로 질병을 도출한 결과는 표 3과 같다.

Table 3. Disease inference results of FCM technique

Diseases	Possibility
Glaucoma	95.24
Dry eye syndrome	66.67
Epiphora	66.67
Cornea wound	28.67
Corneal ulcer	16.26
Scleritis	5.67
Cataract	5.56

표 3에서와 같이 각 질병에는 선택된 증상들이 질병의 일부 증상에 포함되어 있어 가능성들이 95.24에서 5.56으로 나타나는 것을 확인할 수 있다. 이 가능성 값들은 선택된 증상을 기반으로 각 질병에서 나타날 수 있는 증상들의 소속도를 모두 합한 값이다.

IV. Conclusions

전문적인 수의학 지식이 부족한 일반인들이 어떠한 반려동물의 행동이나 몸의 상태 변화를 잘 알지 못하여 자신의 반려동물이 질병에 걸렸다는 것을 인지하지 못하거나 인지하더라도 어떻게 대처해야 할지 모르는 경우가 대부분이다. 이 경우에 대수롭지 않게 생각하여 가축병원에 데리고 가지 않아 간단하게 치료 가능한 질병임에도 불구하고 시기를 놓쳐 반려동물이 큰 고통을 겪게 되고 심각하면

생명이 위험한 상황이 발생할 수도 있다. 따라서 이러한 문제점을 개선하기 위해 가정에서 많이 기르는 반려견에 대해 반려견 질병에 대한 전문적인 수의학 지식이 부족한 일반인들도 자신의 반려견 건강 상태를 확인할 수 있는 반려견 건강 진단 시스템을 본 논문에서 제안하였다.

제안된 반려견 진단 시스템은 증상과 질병에 대한 데이터베이스를 구축하고 질병에 대한 증상들을 FCM 기법을 적용하여 분류하였다. 클러스터링하여 분류된 소속도를 을 기반으로 사용자가 선택한 증상들의 소속도가 가장 높은 질병들을 반려견의 증상과 가까운 질병으로 추론하였다. 그리고 선택한 증상들의 개수와 선택된 증상들이 포함된 질병들의 가능성 값을 구하여 내림차순으로 정렬하여 반려견의 증상과 가장 가까운 상위 질병 3가지를 도출하였다.

그러나 실험에서 알 수 있듯이 해당 질병에 대한 증상 수가 적게 선택되거나 질병과 관련이 매우 적은 증상이 많이 선택되는 경우에는 질병 도출에 대한 정확성이 낮아지는 것을 실험을 통하여 확인하였다.

따라서 향후 연구 방향에서는 반려견 질병에서 대표 증상과 일반 증상의 비율을 조정하여 학습하는 방법을 연구할 것이다. FCM 알고리즘은 소속도를 이용하여 클러스터링 한 후에 소속도를 기반으로 질병을 도출하기 때문에 선택된 증상이 다양한 질병들의 대표 증상의 일부분이거나 다양한 질병들의 일반 증상이 되는 경우에는 진단의 정확성이 낮아지는 문제점이 발생하였다. 따라서 질병에 대한 증상 분류 성능을 개선하기 위하여 퍼지 계층적 클러스터링 기법을 연구하여 대표 증상과 일반 증상을 효과적으로 분류하여 진단의 결과에 대한 정확성을 높일 것이고 진단 결과에 대한 검증 방법도 수의사를 통하여 통계학적으로 진단 결과를 검증하고 분석할 것이다.

ACKNOWLEDGEMENT

Following are results of a study on the "Leaders in INdustry-university Cooperation +" Project, supported by the Ministry of Education and National Research Foundation of Korea.

REFERENCES

[1] M. Martini, et al., "A surveillance system for diseases of companion animals in the Veneto region(Italy)," Scientificand Technical

- Review of the Office International des Epizooties, Vol. 36, No. 3, pp. 1007-1014, Dec. 2017. DOI: 10.1016/j.ijid.2016.11.293
- [2] M. Y. Munirah, et al., "Design and development of online dog diseases diagnosing system," *International Journal of Information and Education Technology*, Vol. 6, No. 11, pp. 913-916, Nov. 2016. DOI: 10.7763/ijiet.2016.v6.816
- [3] G O. Dan, B. C. David, D. M. Paul, C. T. Peter, C. B. Dave , "Approaches to canine health surveillance," *Canine genetics and epidemiology*, Vol. 1, No. 2, pp. 2-10, April 2014. DOI: 10.1186/2052-6687-1-2
- [4] Y. Eui Hong, *Knowledge of pet diseases and how to tame them*, Yuwon publisher, 2001.
- [5] A. Sheshasayee, P. Sharmila, "Comparative Study of Fuzzy C Means and K Means Algorithm for Requirements Clustering," *Indian Journal of Science and Technology*, Vol. 7, No. 6, pp. 853-857, June 2014. DOI: 10.17485/ijst/2014/v7i6.9
- [6] K. B. Kim, D. H. Song, "Intelligent Automatic Extraction of Canine Cataract Object with Dynamic Controlled Fuzzy C-Means based Quantization," *International Journal of Electrical & Computer Engineering*, Vol. 8, Issue 2, pp. 666-672, Apr. 2018. DOI: 10.11591/ijece.v8i2.pp666-672
- [7] K. B. Kim, D. H. Song, S. -S. Yun, "Automatic Extraction of Blood Flow Area in Brachial Artery for Suspicious Hypertension Patients from Color Doppler Sonography with Fuzzy C-Means Clustering," *Journal of information and communication convergence engineering*, Vol. 16, No. 4, pp. 258-283, Dec. 2018. DOI: 10.6109/jicce.2018.16.4.258

Authors



Kwang-Baek Kim received his M.S. and Ph.D. degrees from the Department of Computer Science, Pusan National University, Busan, Korea, in 1993 and 1999, respectively. Dr. Kim is a professor at the Department of

Artificial Intelligence, Silla University, Korea in 2021. He is currently an associate editor for *Journal of Intelligence and Information Systems*. His research interests include fuzzy hierarchical clustering and machine learning, data sciences, data mining, image processing.