

FCM 클러스터링 알고리즘과 퍼지 결정트리를 이용한 상황인식 정보 서비스

양석환[†], 정목동^{**}

요 약

FCM 클러스터링 알고리즘은 대표적인 분할기반 군집화 알고리즘이며 다양한 분야에서 성공적으로 적용되어 왔다. 그러나 FCM 클러스터링 알고리즘은 잡음 및 지역 데이터에 대한 높은 민감도, 직관적인 결과와 상이한 결과 도출 가능성이 높은 문제, 초기 원형과 클러스터 개수 설정 문제 등이 존재한다. 본 논문에서는 FCM 알고리즘의 결과를 해당 속성의 데이터 축에 사상하여 퍼지구간을 결정하고, 결정된 퍼지구간을 FDT에 적용함으로써 FCM 알고리즘이 가지는 문제 중 잡음 및 데이터에 대한 높은 민감도, 직관적인 결과와 상이한 결과 도출 가능성이 높은 문제를 개선하는 시스템을 제안한다. 또한 실제 교통데이터와 강수량 데이터를 이용한 실험을 통하여 제안 모델과 FCM 클러스터링 알고리즘을 비교한다. 실험 결과를 통해 제안 모델은 잡음 및 데이터에 대한 민감도를 완화시킴으로써 보다 안정적인 결과를 제공하며, FCM 클러스터링 알고리즘을 적용한 시스템보다 직관적인 결과와의 일치율을 높여줌을 알 수 있다.

A Context-Aware Information Service using FCM Clustering Algorithm and Fuzzy Decision Tree

Seokhwan Yang[†], Mokdong Chung^{**}

ABSTRACT

FCM (Fuzzy C-Means) clustering algorithm, a typical split-based clustering algorithm, has been successfully applied to the various fields. Nonetheless, the FCM clustering algorithm has some problems, such as high sensitivity to noise and local data, the different clustering result from the intuitive grasp, and the setting of initial round and the number of clusters. To address these problems, in this paper, we determine fuzzy numbers which project the FCM clustering result on the axis with the specific attribute. And we propose a model that the fuzzy numbers apply to FDT (Fuzzy Decision Tree). This model improves the two problems of FCM clustering algorithm such as elevated sensitivity to data, and the difference of the clustering result from the intuitional decision. And also, this paper compares the effect of the proposed model and the result of FCM clustering algorithm through the experiment using real traffic and rainfall data. The experimental results indicate that the proposed model provides more reliable results by the sensitivity relief for data. And we can see that it has improved on the concordance of FCM clustering result with the intuitive expectation.

Key words: FCM Clustering Algorithm(FCM 클러스터링 알고리즘), Fuzzy Number(퍼지구간), Fuzzy Decision Tree(퍼지 결정트리), Context-Awareness(상황인식)

※ 교신저자(Corresponding Author) : 정목동, 주소 : 부산광역시 남구 용소로 45 부경대학교 컴퓨터공학과 컴퓨터보안 및 인공지능 연구실(608-737), 전화 : 051) 629-6253, FAX : 051) 629-6264, E-mail : mdchung@pknu.ac.kr
접수일 : 2013년 5월 20일, 수정일 : 2013년 6월 28일
완료일 : 2013년 7월 12일

[†] 부경대학교 컴퓨터공학과
(E-mail : seokhwan@pknu.ac.kr)

^{**} 부경대학교 컴퓨터공학과

※ 이 논문은 2010년도 부경대학교의 지원을 받아 수행된 연구임(PK-2010-097).

1. 서 론

유비쿼터스 환경의 기반 요소 중 하나인 센서 네트워크는 온도, 습도, 위치와 같은 정보를 시스템에 제공하는 다양한 센서들로 구성되어 있으며 이러한 센서 정보를 적절하게 가공하여 활용한다면 보다 유연한 서비스 제공이 가능하다.

FCM(Fuzzy C-Means) 클러스터링 알고리즘은 1970년대에 제안된 대표적인 군집화 알고리즘이며 현재까지 다양한 분야에서 성공적으로 적용되어 왔다. 그러나 FCM 클러스터링 알고리즘은 잡음 및 지역 데이터에 대한 민감도가 높고[1-4], 계산된 소속도가 직관적인 결과와 상이한 경우가 많으며[5,6], 초기원형과 클러스터 개수 설정 문제[7,8] 등이 존재한다.

이에 본 논문에서는 FCM 알고리즘의 결과를 해당 속성의 데이터 축에 사상하여 퍼지구간을 결정하고, 결정된 퍼지구간을 FDT(Fuzzy Decision Tree)에 적용함으로써 FCM 알고리즘이 가지는 문제 중 잡음 및 지역 데이터에 대한 높은 민감도, 직관적인 결과와 상이한 결과 도출 가능성이 높은 문제를 개선하는 시스템을 제안한다.

본 논문의 구성은 다음과 같다. 2절에서는 관련 연구 내용과 배경 이론에 대하여 살펴보고, 3절에서는 본 논문에서 제안하는 FCM 알고리즘과 FDT를 이용한 분류 알고리즘을 소개한다. 4절에서는 제안 모델을 이용한 상황인식 서비스 모델을 소개하며 5절에서는 제안모델을 이용한 교통상황 예측 시스템을 구현과 실험결과의 분석내용을 다룬다. 마지막으로 6절에서는 결론 및 향후 연구 방향을 제시한다.

2. 관련 연구 및 배경 이론

2.1 FCM(Fuzzy C-Means) 클러스터링 알고리즘

클러스터링은 주어진 데이터를 미리 정해진 클래스와 비교하여 가장 근접한 클래스로 분류하는 기법이며 패턴인식, 의사결정, 데이터 분석 등에서 가장 핵심적인 작업의 하나이다. FCM 클러스터링 알고리즘은 데이터 점을 소속정도에 의해서 분류하는 퍼지 분할 기법을 사용한 데이터 분류 알고리즘이다. FCM 클러스터링 알고리즘의 소속 함수 U 는 0과 1사이의 값을 가지는 요소들을 가지며 데이터 집합에

대한 소속정도 값의 합은 항상 1이다[9]. FCM 알고리즘의 비용함수는 다음과 같은 형태를 가진다.

$$J(u_{ik}, v_i) = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m (d_{ik})^2, \quad (1 \leq m < \infty, v_i = \{v_{i1}, v_{i2}, \dots, v_{ij}, \dots, v_{iL}\})$$

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

$$v_{ij} = \frac{\sum_{k=1}^n (u_{ik})^m x_{kj}}{\sum_{k=1}^n (u_{ik})^m}, u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{2}{m-1}}}$$

u_{ik} : i 번째 클러스터에 속한 x_k 의 k 번째 데이터의 소속도

v_i : i 번째 클러스터의 중심벡터

m : 분류 공정에서 퍼지성의 양을 제어하는 파라미터. 일반적으로 $m=2$

d_{ik} : 클러스터에 속한 k 번째 데이터 x_k 와 i 번째 클러스터의 중심벡터 v_i 사이의 거리

$J(u_{ik}, v_i)$: FCM 클러스터링 알고리즘의 비용함수

표 1은 FCM 클러스터링 알고리즘을 나타낸다[9].

2.2 퍼지 결정트리(FDT: Fuzzy Decision Tree)[10]

결정트리는 분류규칙을 표현하는 트리로서 비 단말모드에는 비교하는 데이터의 특징이 명시되고, 링크에는 비교조건 또는 특징 값이 부여되며, 단말노드에는 루트노드에서 해당 노드까지의 경로 상에 있는 모든 조건을 만족하는 데이터가 속하는 클래스 값이 부여된다. 결정트리는 명확한 값을 기준으로 특징공간을 분할하므로, 미세한 차이를 가지는 서로 다른 두 데이터를 각각 다른 클래스로 분류할 수 있다. 결정트리에 퍼지함수의 개념을 도입하여, 특징공간을 소속함수를 이용하여 정의한 퍼지 경계면으로 분할하는 방법이 퍼지 결정트리(FDT)이다[11,12]. 표 2는 퍼지 결정트리를 이용한 분류과정을 나타낸다.

2.3 상황인식 컴퓨팅(Context-Aware Computing)

상황인식 컴퓨팅이란 유비쿼터스 환경에서 임의의 애플리케이션이 사용자의 다양한 환경 요소에 대한 상황 정보(Contextual Information)를 감지하여,

표 1. FCM 클러스터링 알고리즘

[Step 1]	클러스터 개수 $c (2 \leq c \leq n)$, 지수 가중치 $m (1 < m < \infty)$, 소속 함수 $U^{(0)}$, 알고리즘 반복회수 $r (r = 0, 1, 2, \dots)$ 설정
[Step 2]	클러스터 중심벡터 계산, $v_{ij} = \sum_{k=1}^n (u_{ik})^m x_{kj} / \sum_{k=1}^n (u_{ik})^m$
[Step 3]	새로운 소속 함수 $U^{(r+1)}$ 계산 $u_{ik}^{(r+1)} = \begin{cases} 1 / \sum_{j=1}^c (d_{ik}^r / d_{jk}^r)^{(2/m)-1} & , \text{ for } I_k = \Phi \\ 0 & , \text{ for all classes } i, i \in \tilde{I}_k \end{cases}$ $I_k = \{i 2 \leq c < n; d_{ik}^{(r)} = 0\}, \tilde{I}_k = \{1, 2, \dots, c\} - I_k, \sum_{i \in I_k} u_{ik}^{(r+1)} = 1$
[Step 4]	$\Delta = \ U^{(r+1)} - U^{(r)}\ = \max_{i,k} u_{ik}^{(r+1)} - u_{ik}^{(r)} $ $\Delta > \epsilon : r = r + 1 \rightarrow [Step 2], \Delta \leq \epsilon : \text{Algorithm Finish}$

표 2. 퍼지 결정트리를 이용한 분류과정

[단계 1]	각 노드에 대한 소속 함수, 속성 벡터를 이용하여 T-norm 계산 1) $z_i = w_i^T x$ (각 노드의 속성 벡터 w_i 의 방향으로 입력 데이터 x 사영) 2) 큰 노드에서 n 번째 단말 노드까지의 T-norm 계산 $T_n = \prod_{i = \sqrt{n} \text{ node}}^{n_{th} \leq \text{af node}} \mu_{S_{dw}}^i(Z_i), (0 < T_n < 1)$
[단계 2]	단말 노드들의 T-norm의 평균 계산. 평균 값이 가장 큰 클래스로 분류 $A_i : i$ 번째 클래스를 갖는 단말 노드들의 T-norm의 평균 $class C = \arg_i \max A_i$

사용자가 이를 이용할 수 있도록 해 주는 컴퓨팅 패러다임이다. 상황인식 컴퓨팅은 1994년 Schilit와 Theimer에 의하여 최초로 논의되었으며, Schilit와 Theimer는 상황(Context)이란 사용 장소나 주변의 사람, 사물의 집합 또는 시간이 지나면서 발생하는 변화 등을 일컫는다고 정의하였다[13]. 또한 Dey는 상황이란 엔티티의 상태를 특징지을 수 있는 모든 정보를 말하며, 여기서 엔티티란 사람, 또는 사용자와 애플리케이션 간의 의사소통에 관계되는 사물을 말한다고 정의하고 있다[14]. 상황인식 컴퓨팅은 유비쿼터스 환경 정보를 잘 활용할 수 있는 헬스케어, 보안 등의 분야에서 많이 활용되고 있다.[15]

3. FCM 클러스터링 알고리즘과 FDT를 이용한 분류 알고리즘

FCM 클러스터링 알고리즘은 유클리드 거리를 기반으로 데이터를 분류하는 알고리즘이다. 유클리드

거리를 기반으로 하는 알고리즘은 잡음과 지역 데이터에 대하여 민감하게 반응하는 특징을 가진다[1-4]. 또한 FCM 클러스터링 알고리즘은 간단하고 효과적인 클러스터링 방법이지만, 구해진 소속도가 직관적인 값과 일치하지 않는 경우가 있으며[5,6], 잡음이 많은 환경에서는 정확한 소속도를 구할 수 없다는 문제점을 가진다[16]. 이에 본 논문에서는 FCM 클러스터링 알고리즘을 통하여 도출된 클러스터 영역을 특징 축으로 사상하여 퍼지구간을 결정하고, 결정된 퍼지구간을 FDT에 적용하여 데이터를 분류하는 방법을 제안한다.

FCM 클러스터링 알고리즘을 수행하면 중심벡터와 소속함수가 도출되며, 각 클러스터의 중심벡터와 영역을 이용하여 퍼지구간을 결정한다. 퍼지 규칙을 생성하기 위한 퍼지구간은 삼각형 퍼지수를 사용하며, 삼각형 퍼지수의 꼭지점은 FCM 클러스터링의 수행결과로 도출된 중심벡터의 값과 해당 클러스터가 사상된 축의 최소값, 최대값을 선택한다. 그림 1은

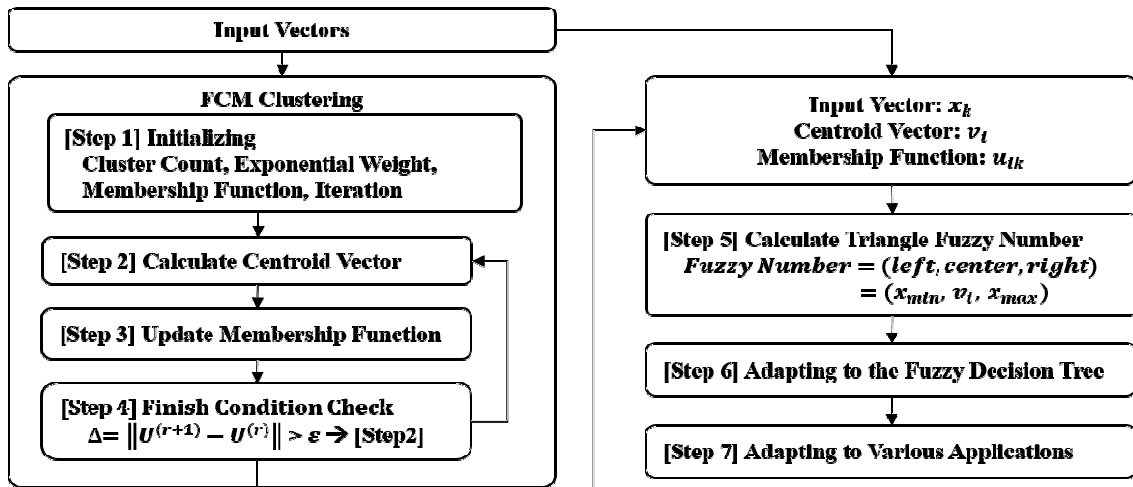


그림 1. 제안 모델의 퍼지구간 생성과정

제안 모델의 퍼지구간 생성과정을 보여준다. 표 3은 제안 모델의 알고리즘을 보여준다.

4. 제안모델을 이용한 상황인식 서비스 모델

4절에서는 FCM과 FDT를 이용한 분류 알고리즘을 실제 시스템에 적용하기 위한 상황인식 정보 서비스 모델을 제안한다. 상황인식 정보서비스 모델은 USN(Ubiquitous Sensor Network)의 센싱 데이터 및 측정 가능한 상황정보의 수집 및 전송을 수행하는 입력부인 Input Interface, 입력된 정보를 이용하여 분류결과를 도출하는 Context-Aware Engine, 도출된 분류결과와 기존의 제어 규칙들을 결합, 참조하여 시스템 제어 규칙을 선택, 수행하는 Control Rule Manager, 모델을 통하여 생성된 규칙과 정보를 외부 시스템으로 제공하기 위한 Output Interface로 구성된다. 상황인식 엔진에서는 Input Interface를 통하여 입력된 USN 데이터와 측정된 상황정보를 제안 모델을 이용하여 분류하고 퍼지 구간을 생성한다. 그림 2는 제안모델을 이용한 상황인식 서비스 모델의 구조를 보여준다. 이 모델은 입력단의 영역을 변경함으로써 다양한 응용프로그램을 구성할 수 있는 프레임워크 개념의 구조이다. 본 논문에서는 5절에서 교통 상황 예측 시스템에 응용하고 있지만 헬스케어나 제조분야 등 다양한 영역으로 적용 가능하며 또한 프레임워크의 기본 API는 재사용할 수 있다. 실제로 연구 과제 수행 중 본 논문에서 제안한 상황인식 정보 서비스 모델을 응용하여 제조업체의 POP-MES 방식의 제조장비 상태 정보를 상황정보로 입력받아 장비

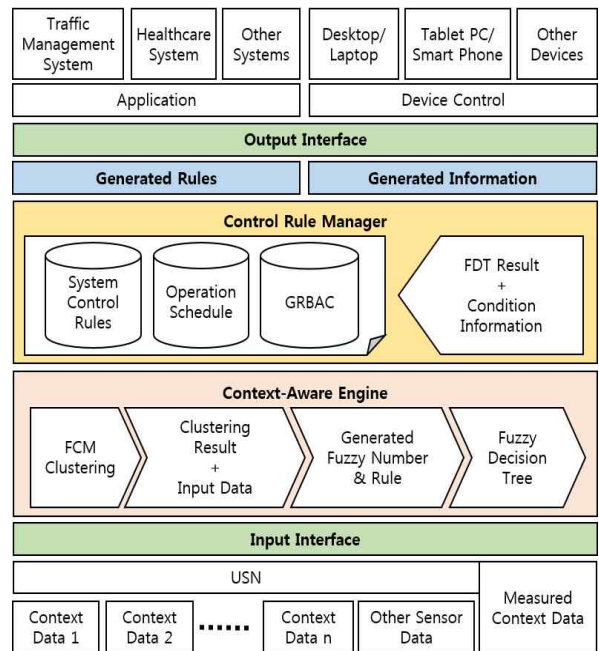


그림 2. 제안 상황인식 정보서비스모델의 구조

피로도 및 위험도 분석에 활용하였으며 API의 70% 가량을 재활용 할 수 있었다.

4.1 Input Interface

Input Interface는 센서네트워크의 다양한 센서들이 제공하는 상황정보를 수집하는 역할을 수행한다. 센서 네트워크는 온도, 습도, 조도, 풍량 등의 환경 데이터, 시간, 위치 및 기타 측정 가능한 상황 데이터를 수집하는 다양한 센서로 구성되어 있다. Input Interface에서 수집한 데이터는 Context-Aware

표 3. 제안 모델의 알고리즘

```

Lfunction SystemClustering(){
    set sensor_weight[x];
    data[x] = inputMethod(); // 센서정보 입력
    res[x] = FCM(data[x]);
    fuzzy_number[] = GenerateFuzzyNumber(res[x], data[x]);
    cl[x] = FuzzyRule(fuzzy_number[c], data[x]);
    congestion_level = FDT(cl[x]);
    executeInfoService(congestion_level);
}

function FCM(data[x]){
    set count_cluster(), exp_weight(), membership_function(), count_repeat();
    while(i < x){ // 센서의 종류 수만큼 수행
        res[i] = getCenterOfCluster(data[x]);
        membership_function = updateMembershipFunction(res[x]);
        if(delta <= e) break;
        else repeat++;
    }
    return res[x];
}

function GenerateFuzzyNumber(res[x], data[x]){
    while(i < clusters){ // 클러스터 수만큼 수행
        while(j < dimension){ // 입력데이터의 차원 수만큼 수행
            fuzzy_number[i][j] = (data[x][j].min, c_vector[j], data[x][j].max);
        }
    }
}

function FDT(cl[x]){
    calculate Tnorm[x];
    Ai = sum(Tnorm[x])/x;
    class = arg(max(Ai));
}

function executeInfoService(congestion_level){ // 다양한 정보서비스 제공
    switch(congestion_level){
        case 1: CongestionType1;
        ...
    }
}

```

Engine으로 전달된다.

4.2 Context-Aware Engine

Context-Aware Engine은 제안모델이 적용된 상황 인식 엔진으로 FCM 클러스터링알고리즘을 이용한 퍼지구간 결정기능과 FDT로 구성되어 있다. Input Interface에서 전달받은 상황 데이터는 제안모델을 통하여 퍼지구간과 분류결과를 계산한다. Con-

text-Aware Engine에서 계산된 분류결과는 상황 데이터와 함께 Control Rule Manager로 전달된다.

4.3 Control Rule Manager

Control Rule Manager는 제안모델이 적용될 시스템에서 실제적인 시스템의 제어를 수행하기 위한 정책을 결정하는 모듈이다. 일반적인 시스템은 작업 스케줄, 상황에 따른 접근규칙, 하드웨어 및 소프트웨어

어에 대한 제어규칙과 같은 제어정책에 대한 정보를 보유하고 있다. Control Rule Manager는 Context-Aware Engine으로부터 전달받은 상황 데이터와 분류결과 데이터를 제어 명령 결정을 위한 입력 데이터로 사용함으로써 시스템의 제어를 위한 Rule을 선택하고 수행한다. 시스템이 제공하는 서비스의 종류에 따라 선택된 Rule과 Context-Aware Engine으로부터 전달받은 상황 데이터를 타 시스템에 제공하게 된다.

4.4 Output Interface

Control Rule Manager에서 제공되는 Rule과 상황 데이터는 Output Interface를 통하여 응용 프로그램 또는 장비제어를 위하여 타 시스템에 제공되게 된다. 다양한 시스템의 지원을 위하여 Output Interface는 추상화 기법을 적용하여 설계되고 서비스를 제공받는 시스템은 추상화 메소드를 통하여 필요한 정보 및 제어신호를 수신하게 된다.

5. 구현 및 평가

본 논문에서는 FCM 클러스터링 알고리즘을 이용하여 입력 데이터로부터 퍼지구간을 계산하고, 계산된 퍼지구간을 FDT에 적용하여 도출된 최종 분류결과를 이용하는 상황인식 정보서비스 모델을 제안하였다. 제안모델의 특징과 성능을 확인하기 위하여 실제 측정 데이터를 이용한 교통상황 예측 시스템을 구현하고, 구현된 시스템을 통하여 기존의 FCM 클러스터링만을 사용한 모델과 제안모델의 비교결과를 제시한다.

실험을 위한 프로그램은 Intel i5-2400 3.1GHz CPU, 4GB Ram이 장착된 IBM 호환 PC에서 C#을 이용하여 구현하였으며 테스트 운영체제는 Microsoft Windows7(64bit)이다.

본 실험에서는 제안모델에 입력되는 상황정보로서 특정 구간에서의 교통량, 차량통행시간, 강수량을 이용하였다. 또한 강수량의 변화에 따른 LOS(Level Of Service)[17]를 제시하고, 제안모델이 제시하는 결과와 기존의 FCM 클러스터링 모델이 제시하는 결과를 비교한다.

실험 데이터는 2011년 6월~7월 사이의 20일에 대하여 측정된 대전지역 교통량, 통행시간 및 강수량

데이터(3개 항목, 1,920건)를 사용하였고, 강수량에 대한 차량운행 속도 및 교통량의 변화를 이용하여 도로교통 서비스 수준을 평가하는 LOS를 계산하였다. 교통량과 통행시간데이터는 1일에 96건으로 15분 주기로 측정하였다. 강수량 데이터는 기상청에서 제공한 해당지역의 실제 측정 데이터를 사용하였으며, 기상청의 데이터는 1시간 단위로 제공되므로 교통량 데이터 4건당 1건으로 동일하게 적용하였다. 각 데이터의 단위는 교통량은 15분간 해당 구간을 지나간 차량의 대수, 통행시간은 해당 구간을 지나가는데 소요된 시간(분), 그리고 강수량은 시간 당 강수량(mm)를 사용하였으며, 실험 데이터는 각 데이터를 0~1 사이의 값으로 정규화 시켜서 사용하였다. 강수량의 정규화는 2004년 5월 개정된 기상법시행령(대통령령 제19555호)과 예보업무규정(기상청훈령 제486호)에 의해 규정된 호우경보 기준인 1시간 강수량 12.5mm를 1로 설정하고 정규화를 수행하였다.

실험은 전체 데이터 1920건(20일분)을 1일 단위로 나누어 총 20개의 서브셋으로 결정하고, 19개의 서브셋은 훈련데이터로, 그리고 1개의 서브셋은 실험데이터로 적용하여 총 20번간의 교차 실험을 수행하여 그 평균값을 계산하여 검증하는 20-Fold Cross Validation 방법을 적용하였다. 20일간의 전체 데이터 중에서 70%인 14일은 강우기록이 있는 날이며, 30%인 6일은 강우기록이 없는 날이다. 20-Fold Cross Validation 방법을 통해 훈련을 종료한 후, 훈련에 사용된 1920건의 데이터 이외의 기간에 측정된 강우기록이 있는 날과 없는 날의 2일간의 데이터 192건을 사용하여, 강수 발생의 시점과 그 시점을 전후하여 분류된 LOS의 결과에 어떤 변화가 일어나는지를 확인하였다. 그림 3은 구현된 프로그램에서 FCM 클러스터링과 제안모델의 분류실험 화면을 보여준다.

표 4는 고속도로 기본구간의 서비스 수준(Level Of Service)에 대한 지표표를 보여준다.

표 4에서 LOS의 평가지표는 고속도로에서의 이상적인 조건일 때를 기준으로 측정된 것이므로 일반도로에서의 결과와는 정확하게 맞지 않으며, 평가 기준으로서의 의미를 가진다. 본 논문에서는 제안모델의 결과로 도출되는 클러스터링 등급을 설명의 편의를 위해서 LOS의 평가지표에 대응시켜서 사용하며, 등급 명은 LOS로 통일하여 사용하기로 한다. 즉 표 4에서 보듯이 LOS의 A~F등급은 제안모델의 1~6

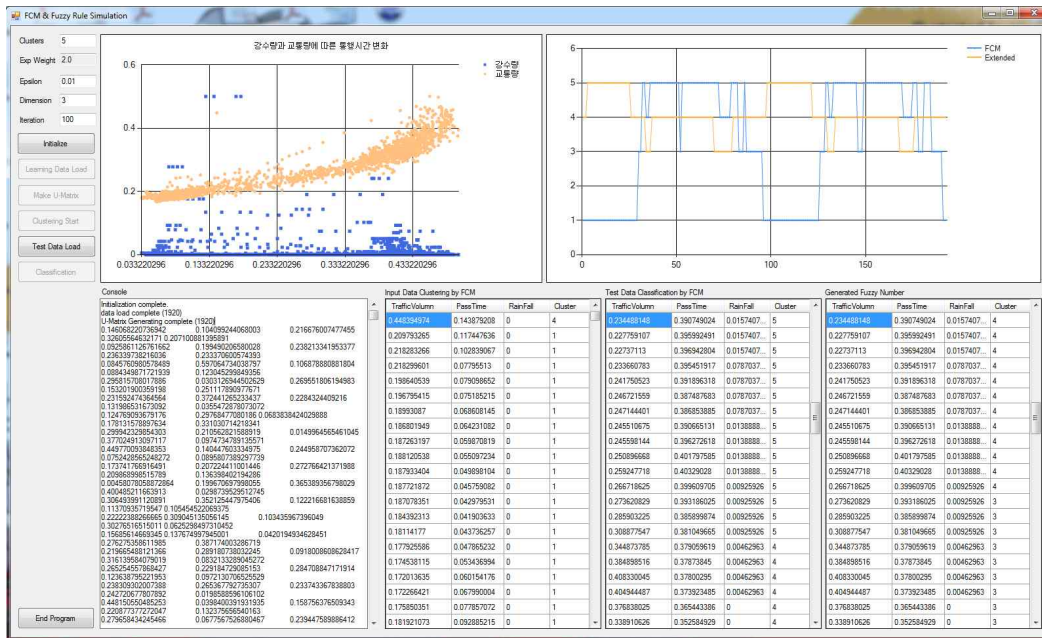


그림 3. 제안모델의 교통상황 분류 실험 화면

표 4. 고속도로 기본구간의 서비스 수준(Level Of Service) [17]

제안모델 등급	LOS	밀도 (pc/km/차로)	설계속도 120 kph		설계속도 100 kph		설계속도 80 kph	
			v/c	MSF ¹⁾	v/c	MSF ¹⁾	v/c	MSF ¹⁾
1	A	≤ 6	≤ 0.30	≤ 700	≤ 0.27	≤ 600	≤ 0.25	≤ 500
2	B	≤ 10	≤ 0.50	≤ 1150	≤ 0.45	≤ 1000	≤ 0.40	≤ 800
3	C	≤ 14	≤ 0.65	≤ 1500	≤ 0.61	≤ 1350	≤ 0.58	≤ 1150
4	D	≤ 19	≤ 0.83	≤ 1900	≤ 0.80	≤ 1750	≤ 0.75	≤ 1500
5	E	≤ 28	≤ 1.00	≤ 2300	≤ 1.00	≤ 2200	≤ 1.00	≤ 2000
6	F	> 28	2)	2)	2)	2)	2)	2)

주: 1) 이상적인 도로 및 교통조건 때의 값임, 2) 불안정류로서 변동이 심함

등급에 대응한다.

표 5는 FCM 클러스터링과 제안모델을 사용하여 실험한 결과를 각각 보여주고 있다. 그러나 이 표에서는 전체 비교 데이터가 아닌 주요 변화를 나타내는 일부 데이터를 제시하였으며 일련번호는 측정당일의 0시부터 24시까지를 15분 단위로 구분하여 2일간 총 192번까지 부여하였다. 통행시간/교통량 데이터는 차량 1대와 관계된 통행시간을 보여준다. FCM 클러스터링의 결과와 제안모델의 LOS 등급은 소통이 원활할수록 1에 가까워지고 정체가 심화될수록 증가하며 최대 값은 6이다.

표 5에서 제시한 데이터에서는 FCM 클러스터링의 결과에서 5번의 LOS 등급의 변화가 발생하였고, 제안 모델에서는 2번의 LOS 등급 변화가 발생하였

다. 표 6은 FCM 클러스터링 결과의 LOS 등급 변화와 제안모델의 LOS 등급 변화를 보여준다. 표 6에서는 표 5에서 제시한 데이터를 차량 당 통행시간을 계산하여 통행시간의 증감을 기준으로 LOS 등급 변화의 방향성을 분석하였다.

표 6에서 FCM 클러스터링의 결과를 보면 (51→52)에서 차량 당 통행시간(통행시간/(교통량*15분))의 값은 0.1035에서 0.1040로 증가하였는데 LOS는 오히려 5에서 3으로 변경되어 소통이 더욱 원활한 것으로 판정되었다(불일치). (52→53)에서는 차량 당 통행시간이 0.1028로 다시 감소하였으나 LOS는 3에서 5로 변경되었다(불일치). 마찬가지로 (82→83)에서는 통행시간은 증가하였으나 LOS는 감소한 것으로 나타남을 알 수 있다(불일치). 그러나 제안모델에

표 5. FCM 클러스터링 알고리즘과 제안모델의 분류결과 비교

번호	A.교통량 (대/15분,구간)	B.통행시간 (분/구간)	C.강수량 (mm/15분)	통행시간/교통량 (B/(A*15))	FCM (LOS등급)	제안모델 (LOS등급)
51	0.208332	0.32133	0.002778	0.103518956	5	4
52	0.209121	0.315178	0.002778	0.104020491	3	4
53	0.215663	0.318695	0	0.102826064	5	4
54	0.225587	0.331755	0	0.100477182	5	4
55	0.235397	0.349462	0	0.098516521	5	4
56	0.240831	0.366581	0	0.098041857	5	4
57	0.240458	0.380685	0.015741	0.098971234	5	4
58	0.234488	0.390749	0.015741	0.101476808	5	4
59	0.227759	0.395992	0.015741	0.105544507	5	4
60	0.227371	0.396943	0.015741	0.111092757	5	4
61	0.233661	0.395452	0.078704	0.115909742	5	4
62	0.241751	0.391896	0.078704	0.116386164	5	4
63	0.246722	0.387488	0.078704	0.112827924	5	4
64	0.247144	0.386854	0.078704	0.108071829	5	4
65	0.245511	0.390665	0.013889	0.104703101	5	4
66	0.245598	0.396273	0.013889	0.104352997	5	4
67	0.250897	0.401798	0.013889	0.10608232	5	4
68	0.259248	0.40329	0.013889	0.10756667	5	4
69	0.266719	0.39961	0.009259	0.106763098	5	4
70	0.273621	0.393186	0.009259	0.103707832	5	3
71	0.285903	0.3859	0.009259	0.099882965	5	3
72	0.308878	0.38105	0.009259	0.095798269	5	3
73	0.344874	0.37906	0.00463	0.089983799	4	3
74	0.384899	0.378738	0.00463	0.082243955	4	3
75	0.40833	0.378003	0.00463	0.073275043	4	3
76	0.404944	0.373923	0.00463	0.065599707	4	3
77	0.376838	0.365443	0	0.061715264	4	3
78	0.338911	0.352585	0	0.061559629	4	3
79	0.303395	0.336409	0	0.064650833	5	3
80	0.269966	0.321804	0	0.069356521	5	4
81	0.240253	0.313708	0	0.07392088	5	4
82	0.215752	0.314238	0	0.079467683	5	4
83	0.197355	0.322451	0	0.087049201	3	4

표 8. FCM 클러스터링 알고리즘과 제안모델의 대표적인 LOS 등급 변화 비교결과

번호	FCM 클러스터링			제안모델				
	차량 당 통행시간	증감	LOS	증감	방향성	LOS	증감	방향성
51 → 52	0.1035	증가	5 → 3	감소	불일치	4 → 4	불변	불변
	0.1040							
52 → 53	0.1040	감소	3 → 5	증가	불일치	4 → 4	불변	불변
	0.1028							
69 → 70	0.1067	감소	5 → 5	불변	불변	4 → 3	감소	일치
	0.1037							
72 → 73	0.0957	감소	5 → 4	감소	일치	3 → 3	불변	불변
	0.0899							
78 → 79	0.0615	증가	4 → 5	증가	일치	3 → 3	불변	불변
	0.0646							
79 → 80	0.0646	증가	5 → 5	불변	불변	3 → 4	증가	일치
	0.0693							
82 → 83	0.0794	증가	5 → 3	감소	불일치	4 → 4	불변	불변
	0.0870							

서는 통행시간의 증감에 대하여 LOS 불일치가 나타나지 않음을 확인할 수 있다.

표 4에서 제시한 LOS의 평가지표에는 강수량의 영향이 고려되지 않았으므로 평가지표의 분류내용이 실험결과와 정확히 일치하지 않지만, 통행시간의 증감과 LOS 변화의 방향성을 비교했을 때 『불일치』 판정인 경우는 잘못된 분류결과라고 판단할 수 있다. 즉 통행시간이 증가하면 LOS도 증가하고 통행시간이 감소하면 LOS도 감소하여야 하는데 표 6을 보면 FCM 클러스터링 분류 결과는 총 3번의 방향성 『불일치』 상태가 나타나고 있지만, 제안모델에서는 2번의 등급변화 모두 방향성의 『일치』 상태를 확인할 수 있다. 표 6을 기준으로 판단할 때 FCM 클러스터링의 결과는 직관적인 판단과 일치하지 않는 경우가 존재함을 확인할 수 있으며, 제안모델의 경우 FCM 클러스터링의 결과보다 개선되었음을 알 수 있다.

또한 표 5, 표 6의 결과 비교에서 FCM 클러스터링의 분류결과는 동일한 구간에서 총 7번의 LOS 변경이 발생하였으나 제안 모델에서는 2번의 LOS 변경이 발생하였음을 확인할 수 있는데, 이는 입력 데이터에 대하여 FCM 클러스터링 보다 제안 모델의 반응 민감도가 낮다는 것을 보여준다. 또한 제안 모델은 입력 데이터에 대한 민감도는 FCM 클러스터링 알고리즘보다 낮지만 전체적으로 안정된 형태의 결과를 보이고 있으며 올바른 결과에 더욱 가까운 결과를 제시하고 있음을 알 수 있다. 실험결과를 통해 제안모델은 FCM 클러스터링이 가지는 문제점인 잡음 및 지역 데이터에 대한 민감도 문제와 계산된 소속도가 직관적인 값과 일치하지 않는 경우가 발생하는 문제를 개선해 줌을 확인할 수 있다.

지금까지의 내용을 정리해 보면 다음과 같다.

① FCM 클러스터링은 3번의 방향성 불일치 판정을 보였으며, 입력데이터에 대하여 직관적으로 알 수 있는 판정과 다른 결론을 도출한 데 반해, 제안 모델은 올바른 결과를 제시함에 따라 보다 안정적인 예측 결과를 도출함을 보여주었다.

② FCM 클러스터링은 입력데이터에 대하여 7번의 LOS 등급 변경을 기록하였으나, 제안 모델은 2번의 LOS 등급 변경을 기록함에 따라 제안 모델이 FCM 클러스터링에 비하여 지역적인 데이터에 대한

민감도가 완화되었음을 보여주고 있다. 즉 제안 모델은 잡음 또는 데이터의 미세한 변화에 대하여 영향을 적게 받는다는 것을 확인할 수 있었다.

6. 결론 및 향후계획

본 논문에서는 FCM 클러스터링 알고리즘을 이용하여 입력 데이터로부터 퍼지구간을 계산하고, 계산된 퍼지구간을 FDT에 적용하여 도출된 최종 분류결과를 적용한 상황인식 정보서비스 모델을 제안하였다. 제안모델의 특징과 성능을 확인하기 위하여 교통상황 예측 시스템을 구현하고, 구현된 시스템을 통하여 기존의 FCM 클러스터링만을 사용한 모델과 제안모델의 비교결과를 제시하였으며, 실험결과를 통해서 제안 모델이 FCM 클러스터링이 가지는 문제점인 잡음 및 지역 데이터에 대한 민감도 문제와 계산된 소속도가 직관적인 값과 일치하지 않는 경우가 발생하는 문제를 개선해 줌을 확인할 수 있었다.

본 실험에서는 실제로 측정된 1,920개의 훈련용 데이터를 사용하였으며, 향후 보다 다양한 실험데이터를 적용할 경우 제안 모델에 대한 더욱 정확한 성능 평가가 이루어질 것으로 기대된다.

향후 연구에서는 사용자 및 상황 별로 상이한 특징을 가진 데이터의 적용이 가능한 차별화된 성능을 제공하는 시스템을 연구할 것이며, 기존의 FCM 클러스터링과 퍼지 제어 시스템이 가지는 한계를 개선할 수 있는 모델을 연구할 계획이다.

참 고 문 헌

- [1] R.N. Dave, "Characterization and Detection of Noise in Clustering," *Pattern Recognition Letters*, Vol. 12, No. 11. pp. 657 - 664, 1991.
- [2] Y. Namkoong, G. Heo, and Y.W. Woo, "An Extension of Possibilistic Fuzzy C-Means with Regularization," *Proc. of the 2010 IEEE Int'l Conf on Fuzzy Systems*, pp. 1-6, 2010.
- [3] Jacek Łęski, "Towards a Robust Fuzzy Clustering," *Fuzzy Sets and Systems*, Vol. 137, No. 2, pp. 215-233, 2003.
- [4] Hesam Izakian and Ajith Abraham, "Fuzzy C-Means and Fuzzy Swarm for Fuzzy Clus-

tering Problem,” *Expert Systems with Applications*, Vol. 38, No. 3, pp. 1835-1838, 2011.

[5] 김태현, 박동철, 정태경, 이윤식, 민수영, “직관적 퍼지 C-평균 모델을 이용한 자기 공명 영상 분할,” *전기전자학회논문지*, 제15권, 제3호, pp. 191-197, 2011.

[6] Dong-Chul Park, “Intuitive Fuzzy C-Means Algorithm,” *IEEE International Symposium on Signal Processing and Information Technology*, pp. 83-88, 2009.

[7] M. Garey, D. Johnson, and H. Witsenhausen, “The Complexity of the Generalized Lloyd-Max Problem,” *IEEE Transactions on Information Theory*, Vol. 28, No. 2, pp. 255-256, 1982.

[8] 허경용, 김광백, “커널 밀도 추정을 이용한 Fuzzy C-Means의 초기화,” *한국해양정보통신학회논문지*, 제15권, 제8호, pp. 1659-1664, 2011.

[9] 오성권, *프로그래밍에 의한 컴퓨터지능*, 내하출판사, 서울, 2002.

[10] R.L.P Chang and T. Pavlidis, “Fuzzy Decision Tree Algorithms,” *Systems, Man and Cybernetics, IEEE Transactions*, Vol. 7, No. 1, pp. 28-35, 1977.

[11] 이우향, 이건명, “특징공간을 사선 분할하는 퍼지 결정트리 유도,” *정보과학회논문지 : 소프트웨어 및 응용*, 제29권, 제3호, pp. 156-166, 2002.

[12] 이건명, “퍼지 데이터에 대한 퍼지 결정트리 기반 분류규칙 마이닝,” *정보과학회논문지 : 소프트웨어 및 응용*, 제28권, 제1호, pp. 64-72, 2001.

[13] B. Schilit, N. Adams, and R. Want, “Context-Aware Computing Applications,” *Proc. of the 1st International Workshop on Mobile Computing Systems and Applications*, pp. 85-90, 1994.

[14] A.K. Dey, *Providing Architectural Support for Building Context-Aware Applications*, Doctoral Dissertation of Georgia Institute of Technology, 2000.

[15] 양석환, 정목동, “상황인식 보안 서비스를 이용한 개선된 접근제어,” *멀티미디어학회논문지*, 제13권, 제1호, pp. 133-142, 2010.

[16] 허경용, 남궁영환, 김성훈, “Regularization을 이용한 Possibilistic Fuzzy C-means의 확장,” *한국컴퓨터정보학회논문지*, 제15권, 제1호, pp. 43-50, 2010.

[17] 도철웅, *교통공학원론(상) 제2개정판*, 청문각, 파주, 2010.



양 석 환

2000년 동서대학교 응용수학과 이학사
 2009년 부경대학교 정보보호학 협동과정 공학석사
 2013년 부경대학교 컴퓨터공학과 박사과정 수료

2000년~2004년 (주) TGS, 산업자원부 부산경남 신발산업 정보화사업 외 다수 프로젝트 개발
 2005년01월~2005년12월 일본 NEC SOFT 개발팀 근무
 2006년03월~2006년08월 (주)나스넷-LG 디스플레이 전산 개발 프로젝트 참여
 2006년09월~2006년12월 부산정보대학 학사관리 시스템 개발 프로젝트 참여
 연구분야 : 신경망(Neural Network), 기계학습(Machine Learning) 등



정 목 동

1981년 경북대학교 컴퓨터공학과 공학사
 1983년 서울대학교 컴퓨터공학과 공학석사
 1990년 서울대학교 컴퓨터공학과 공학박사

1984년~1985년 금성반도체(주) 연구소 연구원
 1985년~1996년 부산외국어대학교 컴퓨터공학과 교수
 1996년~현재 부경대학교 컴퓨터공학과 교수
 1999년~2000년 미국 IOWA 대학교 방문교수
 연구분야 : 컴퓨터응용보안 (Computer Security for Application), 컨텍스트 인지 컴퓨팅(Context Aware Computing), 지능형 에이전트(Intelligent Agent) 등