

2018



4차 산업 관련 인공지능 기반 기술 워크숍



일시 : **2018년 8월 24일** **금요일** 9시 30분~17시

주최 : 한국정보통신학회 지능정보연구회

주관 : 한국정보통신학회, 부산인재평생교육진흥원

장소 : 밸류호텔부산

KIICE
한국정보통신학회

사단
법인 한국정보통신학회

2018



4차 산업 관련 인공지능 기반 기술 워크숍



일시 : **2018년 8월 24일** 금요일 9시 30분~17시

주최 : 한국정보통신학회 지능정보연구회

주관 : 한국정보통신학회, 부산인재평생교육진흥원

장소 : 벨류호텔부산

목차

05 환영사

06 초대의 말씀

07 연사소개

08 프로그램

09 강의내용

▮ 영상 분석 및 처리를 위한 퍼지 학습 알고리즘 · 9
김광백 교수 (신라대학교)

▮ 유전 알고리즘 : 좋은 것이 남는 진화의 법칙 · 49
우영운 교수 (동의대학교)

▮ 의사결정트리를 활용한 분류기법 · 71
김희철 교수 (인제대학교)

한국정보통신학회 지능정보연구회

환영사

한국정보통신학회 회원 및 산업체 전문가 여러분,

현재 우리가 살고 있는 세상은 네트워크(IoT, 5G), 데이터(Cloud, Big Data), 인공지능(기계학습, 알고리즘) 등 지능 정보통신기술(ICT)이 기존 산업과 융합하여 새로운 서비스와 가치를 창출하는 4차 산업혁명 시대이며, 세상의 모든 것들이 빠르게 변화하는 새로운 사회의 도래를 목전에 두고 있습니다.

이러한 급격한 시대 변화의 중심에 한국정보통신학회가 자리하고 있습니다. 우리 학회는 21년이란 짧은 역사에도 불구하고 11회의 국제학술대회와 43회의 국내학술대회를 개최하였고 매달 발간되는 국문지와 분기별 발간되는 영문지가 한국연구재단의 등재지로서 그 권위를 더 하고 있습니다. 또한, 작년에는 영문논문지(JICCE)가 국제 유명인용 색인인 SCOPUS에 등재되는 성과를 이루었습니다.

우리 학회는 이 4차 산업혁명 시대를 맞이하여 IoT, 5G, AI, Big Data, 클라우드 등의 학문적 발전을 도모하고 관련 기술에 종사하는 학계 및 산업체와의 유대관계를 공고히 하고 이를 기반으로 우리나라 정보통신 산업을 선진국 수준으로 발전시켜야 하는 시대적 사명을 지고 있다고 생각합니다.

이에 한국정보통신학회 지능정보연구회 주관으로 개최되는 본 기술 워크숍은 매우 시기 절적하며, 관련 산업체에 계시는 분들 뿐만 아니라 이 분야를 연구하고 관심을 갖는 학생들에게도 인공지능 관련기술의 실무적 know-how와 깊이 있는 이론적 지식을 공유하는 의미 있는 자리가 될 것으로 생각합니다. 금번 기술 워크숍에서의 활발한 의견교환과 지식공유를 통해 우리나라 ICT 산업이 한 단계 더 발전하는 기회가 되길 기대하며, 참석하신 모든 분들의 적극적인 참여를 부탁드립니다.

끝으로 본 기술 워크숍을 준비해주신 지능정보연구회 위원장인 김광백교수님을 비롯한 관계자분들의 노고에 깊은 감사를 드리며, 참석해 주신 모든 분들의 가정에 행복이 가득하시길 기원합니다. 고맙습니다.

2018년 8월

사단법인 한국정보통신학회 회장 **오창헌**

초대의 말씀

4차 산업혁명을 통해 인공지능기술 및 사물인터넷, 빅데이터 등 정보통신기술(ICT)과의 융합을 통해 생산성이 급격히 향상되고, 제품과 서비스가 지능화되면서 경제·사회 전반에 혁신적인 변화가 나타나고 있습니다. 또한 4차 산업혁명은 디지털 혁명으로 요약되는 3차 산업을 기반으로 물질적, 디지털, 생물학적 영역의 경계를 허무는 기술 혁명으로 분석되고 있습니다. 특히 4차 산업 혁명의 기반 기술은 인공지능 기술로서 기계 학습, 심층 신경망, 퍼지 시스템, 유전 알고리즘 등이 실무에서 다양하게 적용되고 있습니다. 그러나 대부분의 현장에서는 기계 학습 알고리즘이나 딥러닝 알고리즘 등의 학습 구조 및 알고리즘의 장단점 분석 없이 오픈소스 등을 적용하고 있는 실정입니다. 특히 딥러닝의 기본 기법인 CNN이나 RNN 등을 TensorFlow를 기반으로 파이썬을 적용하여 다양한 응용 분야에 적용하고 있습니다. 그러나 이들 알고리즘에 대한 학습 구조 및 학습 방법을 이해하여 적용하는 수준은 매우 낮은 편에 있습니다.

따라서 지능 정보 연구회에서는 1차 워크숍으로 상용화 소프트웨어에서 인공지능 기술을 접목하고자 하는 개발자를 위하여 최적화 접근을 위한 유전 알고리즘, 의사결정트리를 활용한 다양한 기능 분류 기법, 빅데이터 기반 특징 분류를 위한 클러스터링 기법 및 개선 방향 등에 대해 강연하고 관련된 프로그래밍 작성 방법과 프로그램 소스를 제공합니다.

본 워크숍은 현장에서 빅데이터 처리나 인공지능 기술이 활용되는 분야에서 오픈 소스나 라이브러리를 사용하는 수준을 벗어나 직접 알고리즘과 구조를 이해하고 개선하여 프로그램을 작성할 수 있는 능력을 향상시키는 수준으로 준비하였습니다.

이번 워크숍을 시작으로 2019년에는 다양한 기계 학습 및 최신 딥러닝을 포함한 심층 신경망 알고리즘들의 장단점을 분석하고 단점들을 개선하는 방법 및 프로그램 작성 방법 등에 대한 워크숍을 계획하고 있습니다.

이번 워크숍이 실무 개발자의 수준을 향상시키는 계기가 되기를 바라며, 개발자 및 연구자들의 적극적인 참여와 성원을 부탁드립니다. 감사합니다.

한국정보통신학회 지능정보연구회 위원장 김광백

연사 소개



김광백 (Kwang Baek Kim) 교수 신라대학교

1999년 : 부산대학교 전자계산학과 졸업(이학박사)
1997년~현재 : 신라대학교 컴퓨터소프트웨어공학부 교수
2013년 : International Journal of Computational Vision and Robotics(SCOPUS), Guest Editor
2013년 : International Journal of Information and Communication Technology(SCOPUS), Guest Editor
2015년~2016년 : Computational Intelligence and Neuroscience(SCIE), Lead Guest Editor
2016년~2017년 : 한국정보통신학회 회장
2014년~현재 : Open Computer Science Journal, Editor.
2012년~현재 : 한국지능정보시스템학회 편집위원
2013년~현재 : International Journal of Intelligent Information Processing(SCOPUS), Editor
2015년~현재 : International Journal of Fuzzy Logic and Intelligent Systems(SCOPUS), Associate Editor
2017년~현재 : IEEE Computational Intelligence Society 이사
논문(2013~2018) : SCIE(1저자 & 교신저자 : 26편), SCOPUS(1저자 & 교신저자 : 28편)
연구분야 : 심층 신경망, 기계 학습, 의료 영상 처리, 퍼지 시스템, 데이터마이닝



우영운 (Young Woon Woo) 교수 동의대학교

1997년 8월 : 연세대학교 전자공학과 (공학박사)
1997년 9년~현재 : 동의대학교 응용소프트웨어공학과 교수
2016년~현재 : 한국정보통신학회 국문지 편집위원장
2018년~현재 : 한국정보통신학회 수석부회장
2017년 5월 : 'AI 이해와 활용' 특강(KTDS)
2017년 8월 : '인공지능의 이해와 활용' 특강(인제대학교 대학원)
논문 및 학술발표(2013~2018) : SCOPUS(4편), 학진등재지(7편), 국제학술발표(5편), 국내학술발표(9편)
역서 : <엑셀로 배우는 인공지능> (제이펍, 2017)
연구 분야 : Artificial Intelligence, Machine Learning, Fuzzy Techniques, Pattern Recognition, Data Mining



김희철 (Hee-Cheol Kim) 교수 인제대학교

2001년 : Stockholm 대학교 수치해석/컴퓨터과학과 졸업(이학박사)
2002년~현재 : 인제대학교 컴퓨터공학부/헬스케어IT학과 교수
2016년~현재 : 인제대학교 일반대학원 디지털항노화헬스케어학과 학과장
2016년~현재 : 경상남도교육청 정보화정책심의위원
2016년~현재 : Journal of Multimedia Information Systems, 편집위원
2017년~현재 : 한국정보통신학회 편집위원
2005년~2010년 : IEEE HealthCom 편집위원
논문 및 연구 : HCI, 인공지능, 의료정보학 분야에서 100여 편 이상의 논문 출간.
<나노섬유 기반 웰니스웨어 시스템 개발, 2009~2014>, <창의융합산업 특성화 인재양성사업, 2016~2021> 외 다수의 연구 및 교육 프로젝트를 연구책임자로 수행
저서 <인간과 컴퓨터의 상호작용>(사이텍미디어, 2006) : 문화관광부 우수학술도서
연구분야 : 인공지능, 기계학습, 빅데이터 마이닝, 디지털 헬스케어, 휴먼 컴퓨터 인터페이스

4차 산업 관련 인공지능 기반 기술 워크숍 프로그램

시간	주제	발표자
09:30 ~ 10:00	등록	
10:00 ~ 12:00	<p>영상 분석 및 처리를 위한 퍼지 학습 알고리즘</p> <ol style="list-style-type: none"> 1. Fuzzy C-Means <ul style="list-style-type: none"> - Fuzzy C-Means Algorithm - Deep FCM 개발 방법 2. Fuzzy Max-Min Neural Networks <ul style="list-style-type: none"> - Fuzzy Max-Min Neural Networks Algorithm 3. Deep 기반 Fuzzy Hybrid Learning Algorithm <ul style="list-style-type: none"> - 학습 구조 - 학습 알고리즘 - 개선 방향 및 개선 방법 	김광백 교수 (신라대학교)
12:00 ~ 13:00	점심시간 (점심 제공)	
13:00 ~ 14:50	<p>유전 알고리즘 : 좋은 것이 남는 진화의 법칙</p> <ol style="list-style-type: none"> 1. 유전 알고리즘의 개요 2. 유전 알고리즘의 처리 기법 3. 유전 알고리즘의 구체적 예 4. 유전 알고리즘의 응용 분야 	우영운 교수 (동의대학교)
14:50 ~ 15:00	Break	
15:00 ~ 16:50	<p>의사결정트리를 활용한 분류기법</p> <ol style="list-style-type: none"> 1. 의사결정트리의 정의 2. 의사결정트리의 특성 3. 의사결정트리 활용 분야 4. 비선형 데이터 분류를 위한 퍼지결정트리 	김희철 교수 (인제대학교)

4차 산업 관련 인공지능 기반 기술 워크숍

영상 분석 및 처리를 위한 퍼지 학습 알고리즘

10:00~12:00

김광백 교수(신라대학교)

영상 분석 및 처리를 위한 퍼지 클러스터링 및 학습 알고리즘

Deep Neural Networks & Medical Imaging LAB
Silla University
Kwang Baek Kim

1. 퍼지 이론

Fuzzy 개념

- 퍼지 이론은 이치 논리가 아니라 다치 논리임.
- 애매한 기준을 수치적으로 값을 표현

Example
 뜨겁다 차갑다(이치 논리)
 -> 1, 0
 매우 뜨겁다 뜨겁다 미지근하다 약간 차갑다 차갑다(다치 논리)
 -> 1, 0.8, 0.6, 0.4, 0.2, 0

애매함을 수치적으로 취급이 가능하도록 하는 이론 → 퍼지 이론

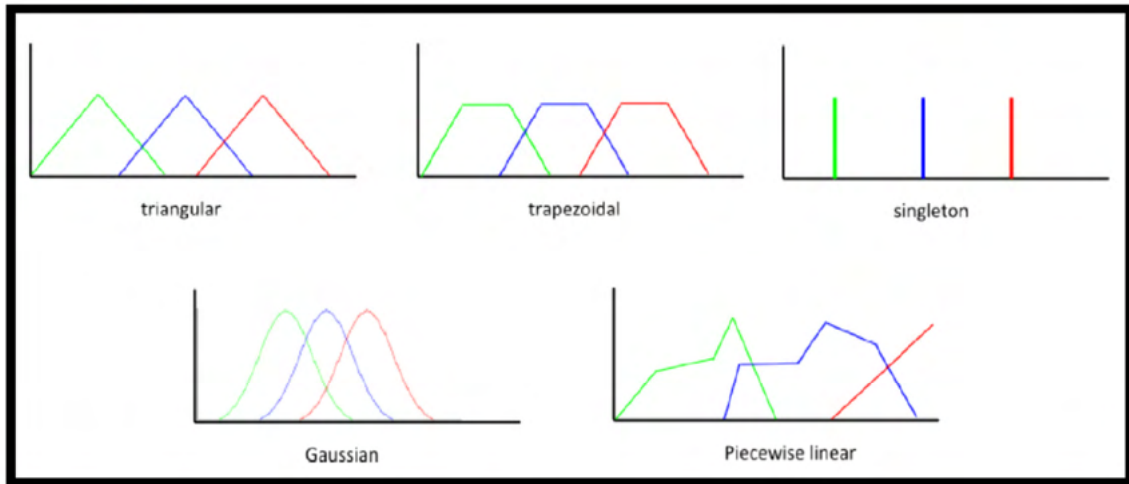
Fuzzy Set

퍼지 집합

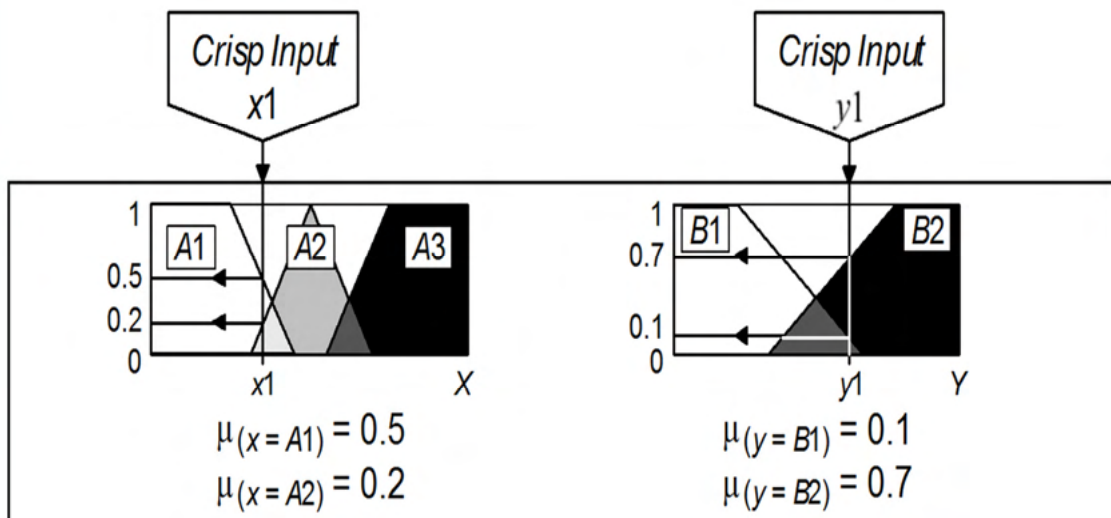
	큰키	중간키	작은키	결론
176cm	0.6	0.4	0	키가 큰 것보다 약간 작다
166cm	0	0.6	0.4	키가 중간 보다 약간 작다
163cm	0	0.3	0.7	키가 작다

Membership Functions

- **Triangular.**
- **Trapezoidal.**
- **Piecewise linear.**
- **Gaussian.**
- **Singleton.**



Fuzzification

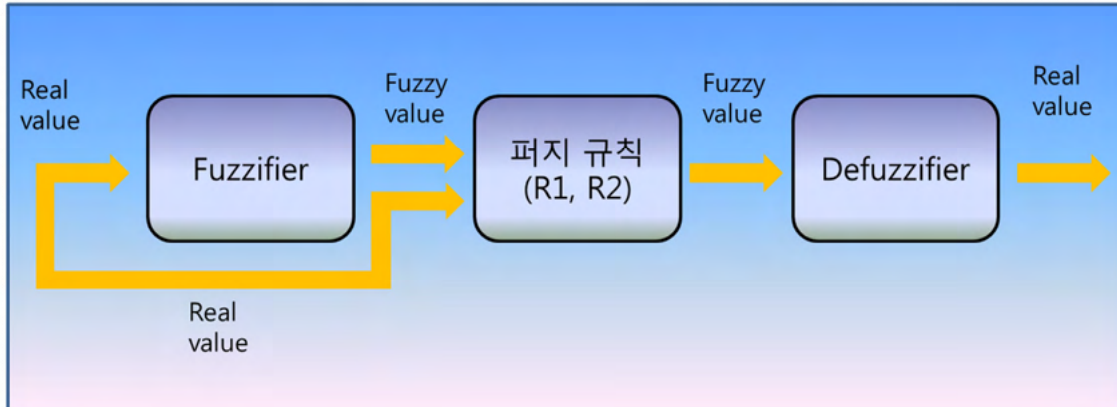


Fuzzy Control Method

신라대학교



If 실온이 [약간 높다] and 습도가 [꽤 높다],
then 에어컨의 스위치를 [강하게] 하시오.



Fuzzy Inference

신라대학교



If x is A1 and y is B1, Then z is C1
If x is A2 and y is B2, Then z is C2

$$R_1 \text{의 적합도 : } W_1 = \mu_{A1}(x_0) \wedge \mu_{B1}(y_0)$$

$$R_2 \text{의 적합도 : } W_2 = \mu_{A2}(x_0) \wedge \mu_{B2}(y_0)$$

$$R_1 \text{의 추론 결과 : } \mu_{c1}(z) = W_1 \wedge \mu_{c1}(z), \forall z \in Z$$

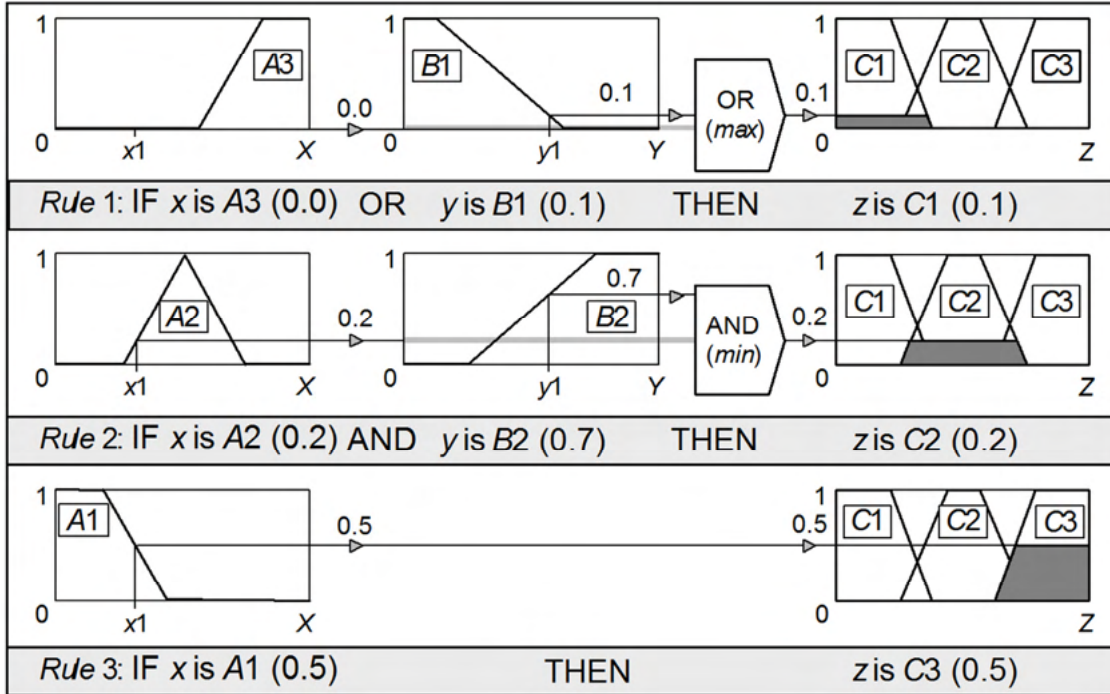
$$R_2 \text{의 추론 결과 : } \mu_{c2}(z) = W_2 \wedge \mu_{c2}(z), \forall z \in Z$$

$$\mu_c(z) = \mu_{c1}(z) \vee \mu_{c2}(z)$$

$$z_0 = \frac{\int u_c(z) \cdot z dz}{\int u_c(z) dz}$$

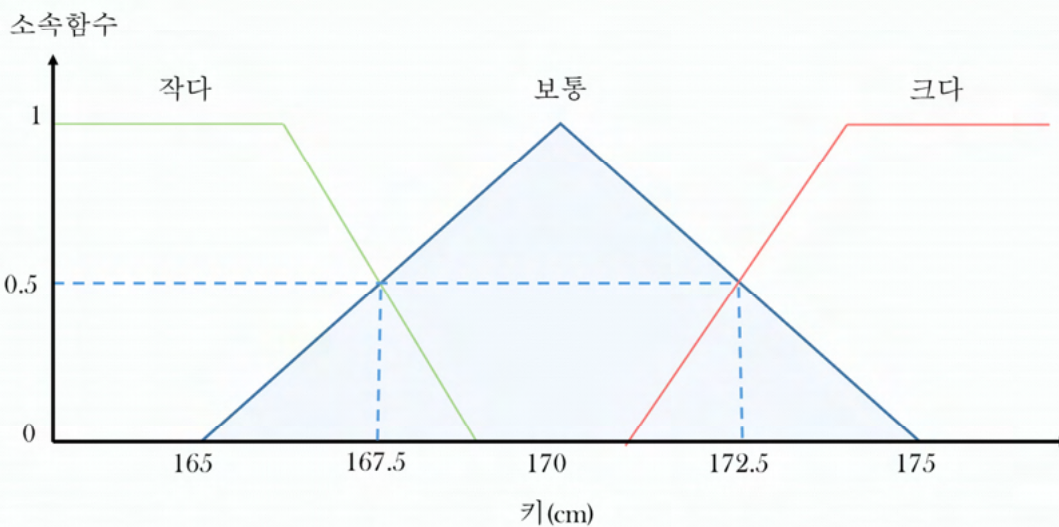
Mamdani-style rule evaluation

신라대학교



Membership Function 예제

신라대학교



무게 중심법

신라대학교



무게중심법은 소속함수의 값과 출력 값의 곱을 모두 더한 것으로 변환하는 것이다.

예를 들어 키가 작은 사람이 하루에 마시는 우유가 0.5L, 보통인 사람이 0.8L, 키가 큰 사람이 1.4L의 우유를 마신다고 가정한다.

어떤 사람의 키가 168.5cm에 소속 함수의 값이 작다@0.3, 보통@0.7이라면 무게중심법에 의해 추론하는 하루의 우유 섭취량은 0.71L이다.

$$\text{우유 섭취량} = \frac{0.3 \times 0.5L + 0.7 \times 0.8L}{0.3 + 0.7} = 0.71L$$

신라대학교



2. Fuzzy Max-Min Neural Networks

지도 학습 Supervised Learning

지도 학습

- 동전 → 동전
- 음식 → 음식

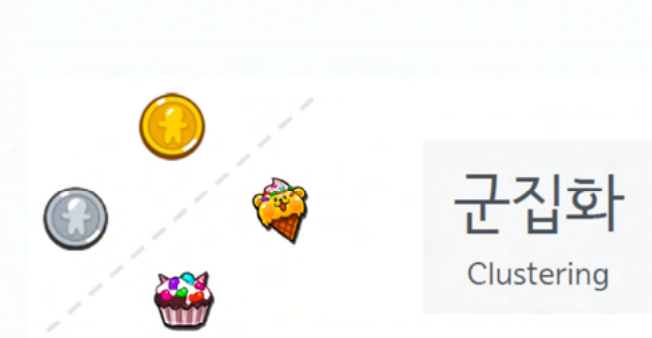
- ? → ?
- ? → ?

분류
Classification

비지도 학습 Unsupervised Learning

비지도 학습

- ? → ?
- ? → ?
- ? → ?
- ? → ?



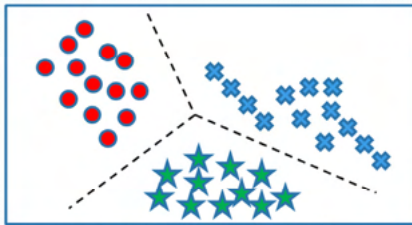
Learning Algorithm : 학습 방법



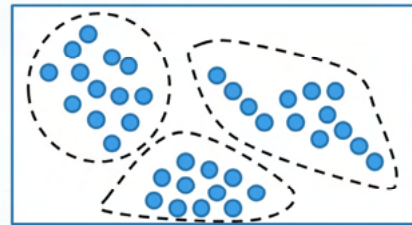
Traditional Programming



Artificial Learning

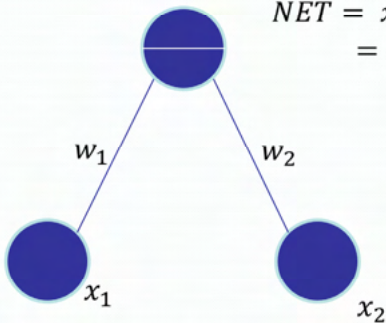


Supervised learning



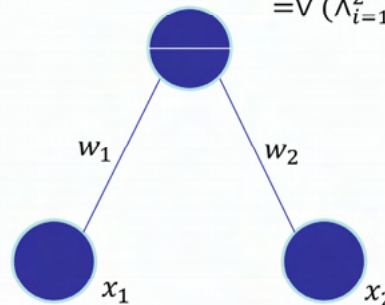
Unsupervised learning

Fuzzy Max-Min Neural Networks



단층 퍼셉트론

$$\begin{aligned}
 & \text{if } NET \geq 1 \text{ Then } Y = 1 \\
 & \text{else } Y = 0 \\
 & NET = x_1 w_1 + x_2 w_2 \\
 & = \sum_{i=1}^2 x_i w_i
 \end{aligned}$$

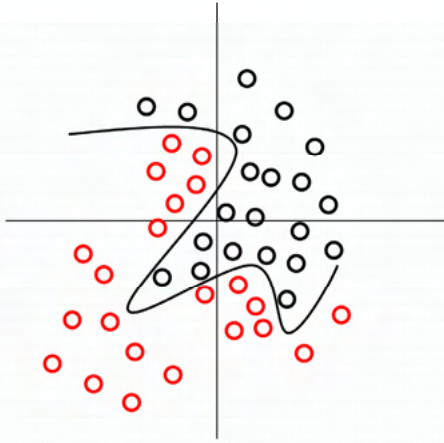


Fuzzy Max-Min Neural Networks

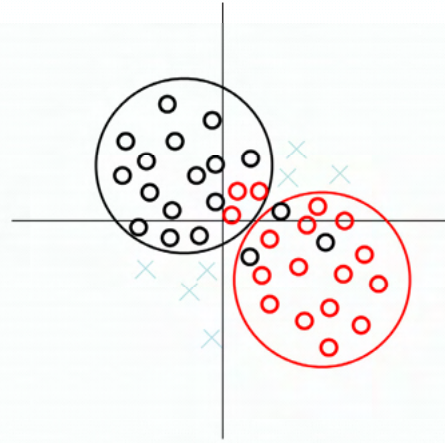
$$\begin{aligned}
 Y &= NET \vee \theta_i \\
 NET &= (x_1 \wedge w_1) \vee (x_2 \wedge w_2) \\
 &= \vee (\wedge_{i=1}^2 (x_i, w_i))
 \end{aligned}$$

지도 학습 & 비지도 학습

신라대학교



지도 학습



비지도 학습

학습단계 분석

신라대학교



학습 단계

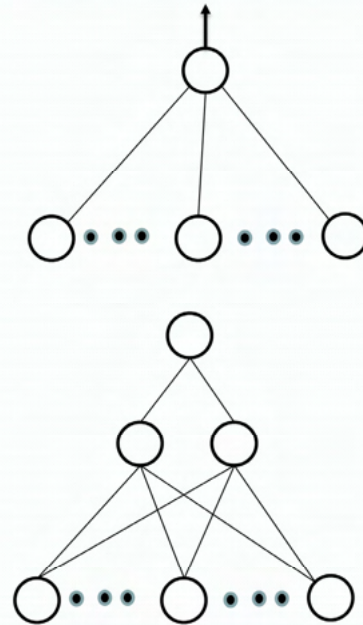
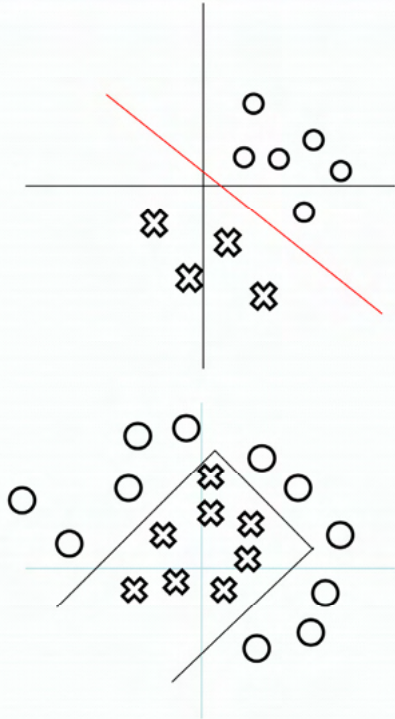
- Fuzzy Max-Min Neural Network, BP, CNN 등과 같은 지도 학습 알고리즘

1. 수렴 단계
 - 학습 초기에 오차가 급격히 감소하는 단계
2. 경쟁 단계
 - 오차의 변화가 거의 없거나 오차 값이 진동하는 단계
3. 우세 단계
 - 경쟁 단계를 벗어나 오차가 급격히 감소하면서 학습되는 단계



학습 구조 기반 분류 방법

신라대학교



퍼지 Max-Min 신경망 알고리즘

신라대학교



- NET는 입력 벡터(x_i)와 연결 가중치(w_i)의 최대-최소 합성 연산에 의해 식(1)과 같이 계산

$$NET = \bigvee_{j=1}^q \{ \bigwedge_{i=1}^p (x_i, w_{ji}) \} \quad \text{식(1)}$$

- 출력 벡터(o_j)는 식(2)와 같이 계산

$$o_j = NET \vee \theta_j \quad \text{식(2)}$$

퍼지 Max-Min 신경망 알고리즘

신라대학교



- 출력 벡터(o_j)와 목표 벡터(t_j)가 동일한 경우에는 연결가중치(w_{ji})와 바이어스항(θ_j)을 변경하지 않음
- 동일하지 않을 경우 식(3)에 의해 연결가중치(w_{ji})와 바이어스항(θ_j)을 변경

$$\begin{aligned}
 \Delta w_{ji}(n+1) &= \Delta w_{ji}(n) + \frac{\partial o_i}{\partial w_{ji}}(t_j - o_j) & \frac{\partial o_j}{\partial w_{ji}} &= 1, \text{ when } o_j = w_{ji}, \\
 & & &= 0 \text{ otherwise.} \\
 \Delta \theta_j(n+1) &= \Delta \theta_j(n) + \frac{\partial o_j}{\partial \theta_j}(t_j - o_j) & \frac{\partial o_j}{\partial \theta_j} &= 1, \text{ when } o_j = \theta_j, \\
 & & &= 0 \text{ otherwise.}
 \end{aligned}
 \tag{3}$$

$$\begin{aligned}
 w_{ji}(n+1) &= w_{ji}(n) + \alpha \Delta w_{ji}(n) + \beta \Delta w_{ji}(n-1) \\
 \theta_j(n+1) &= \theta_j(n) + \alpha \Delta \theta_j(n) + \beta \Delta \theta_j(n-1)
 \end{aligned}$$

퍼지 Max-Min 신경망 알고리즘

신라대학교



종료
조건

총 오차 자승합(TSS)이 오류 한계(ρ)보다 크면 단계 3으로 가고 오류 한계보다 적거나 같으면 학습을 종료

$$TSS = \sum_{p=1}^h \sum_{j=1}^q \frac{1}{2} (t_j^p - o_j^p)^2$$

$TSS > E_{MAX}$ Then Step 3.
 otherwise 학습 종료

3. Fuzzy C-Means Clustering

Clustering

- Clustering
 - Unsupervised 학습
 - 유사성의 개념을 바탕으로 데이터를 몇 개의 그룹으로 분류하는 방법
 - 패턴인식, 문헌검색 등에 널리 응용되고 있음
 - ❖ Cluster : 사전적으로 무리라는 의미로 전산 분야에서는 동일 속성을 갖는 대상들을 하나로 묶은 대상을 의미

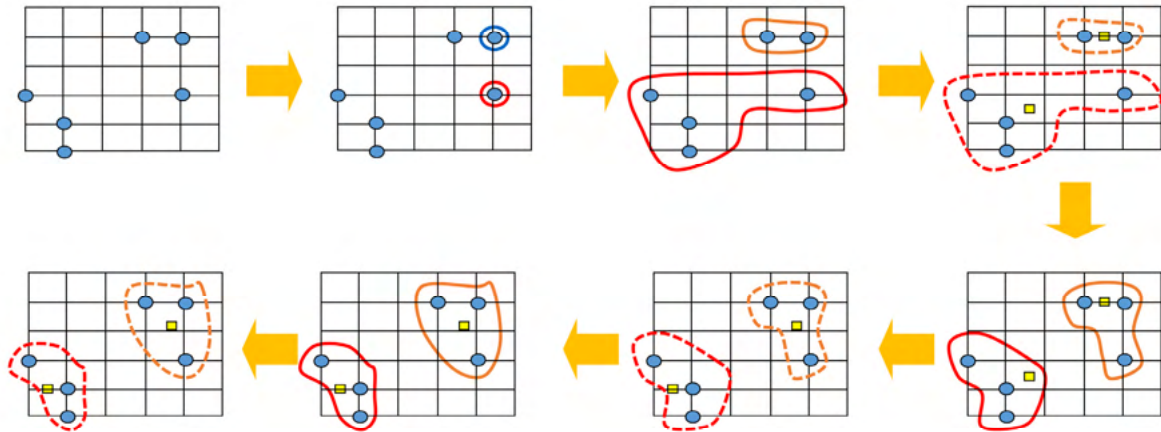


Clustering 과정

신라대학교



• Clustering Algorithm



클러스터링 기법 비교

신라대학교



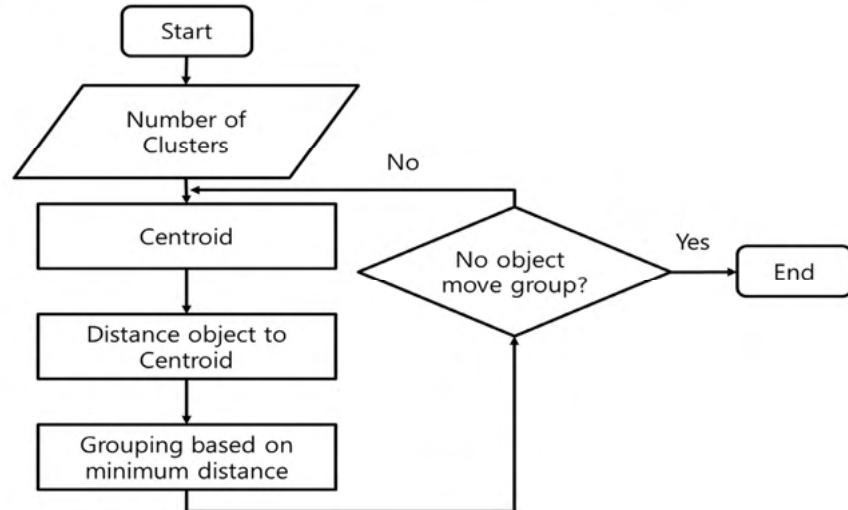
알고리즘	데이터 특징	클러스터 생성	Deep Unsupervised Neural Network 확장
K-Means	정량화된 빅데이터	정적 생성	가능 (비효율적)
PCM	정량화된 빅데이터	정적 생성	가능 (비효율적)
FCM	비정량화된 빅데이터	정적 생성	가능 (효율적)

K-Means 알고리즘

신라대학교



- K-Means 알고리즘의 개념은 패턴들과 그 패턴이 속하는 클러스터의 중심과의 평균 유클리디안(Euclidean)거리를 최소화 하는 것이다.
- K-Means 알고리즘의 성능은 초기 중심을 어떻게 선정하는가에 따라 크게 달라진다.



FCM(Fuzzy C-Means)

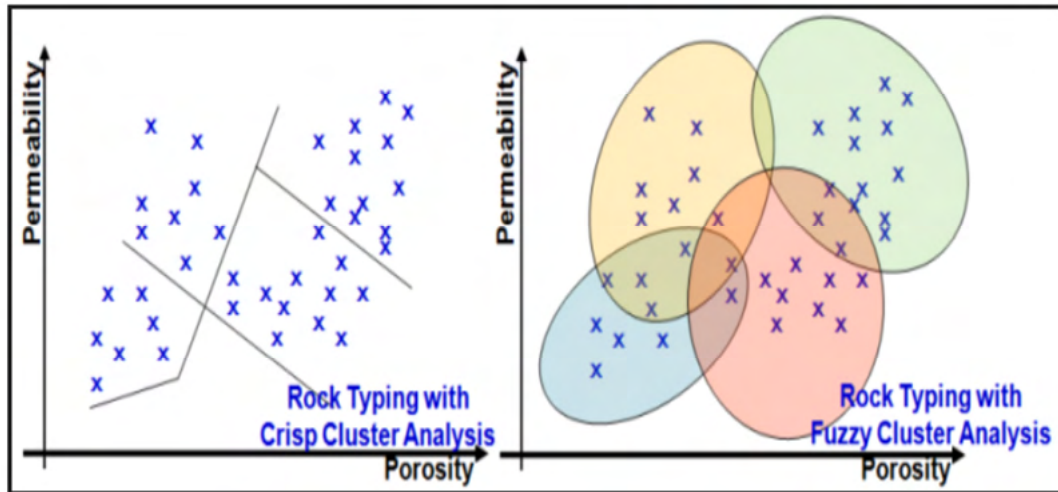
신라대학교



- FCM(Fuzzy C-Means)
 - 각각의 데이터를 각 클러스터에 속하는 소속 정도를 이용하여 데이터 분류하는 알고리즘
 - 유사성을 유클리디안 거리값을 이용하여 측정
 - HCM 알고리즘은 하나의 데이터가 하나의 클러스터에만 속할 수 있지만 FCM은 퍼지 소속정도를 이용하여 2개 이상의 클러스터에도 속할 수 있음
- ❖EM알고리즘 : Expectation과 Maximization단계를 반복하여 최적해를 찾는 알고리즘
- ❖HCM(K-Means) : 유클리디안 거리를 이용하여 K개의 그룹으로 분류하는 클러스터링 알고리즘

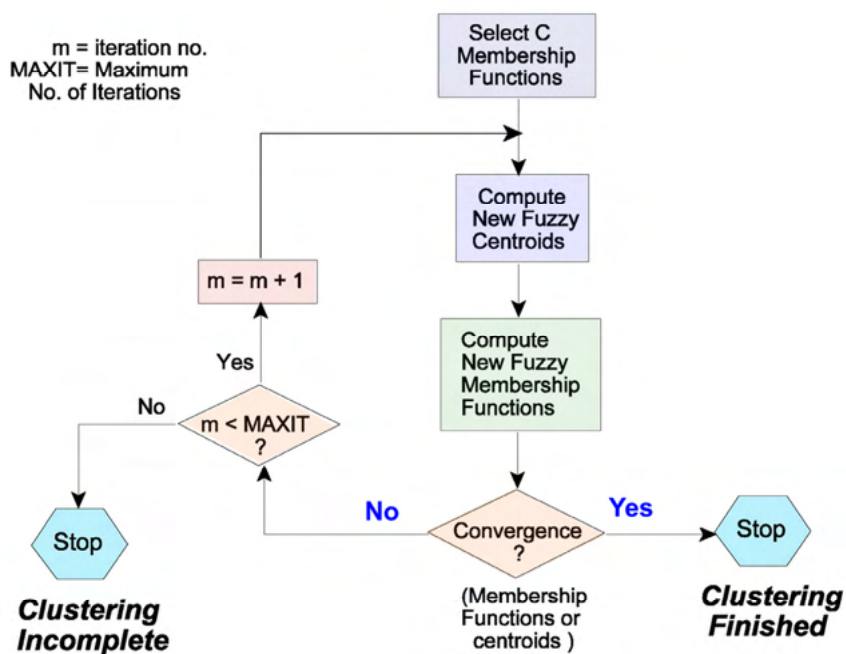
Crisp Clustering & Fuzzy Clustering

신라대학교

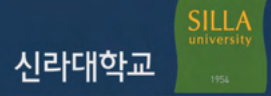


Fuzzy C-Means Clustering Algorithm Flow Diagram

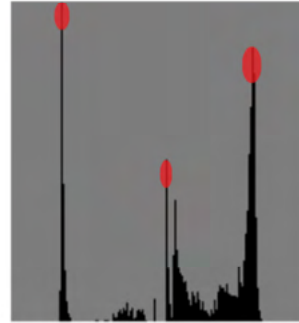
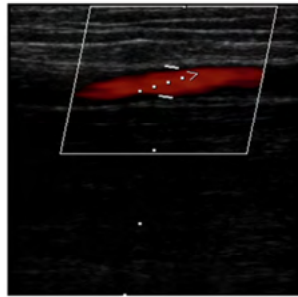
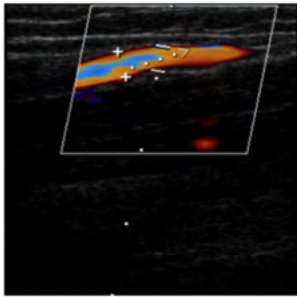
신라대학교



FCM 기반 색조 도플러 초음파 영상에서 상완 동맥의 혈류 영역 추출

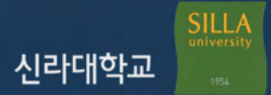


클러스터 개수 정의

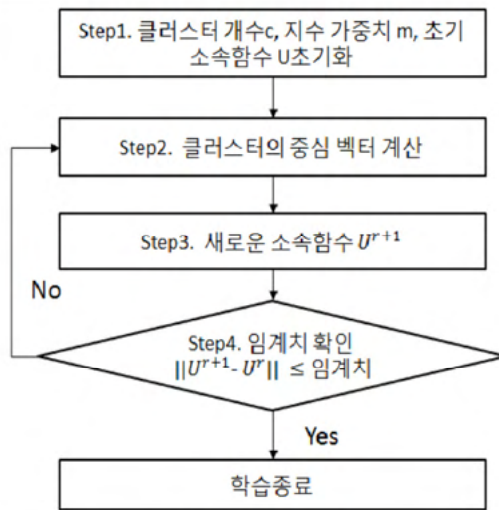


히스토그램 분석을 통한 봉우리 지점 개수 추출

Fuzzy C-Means Alorithm



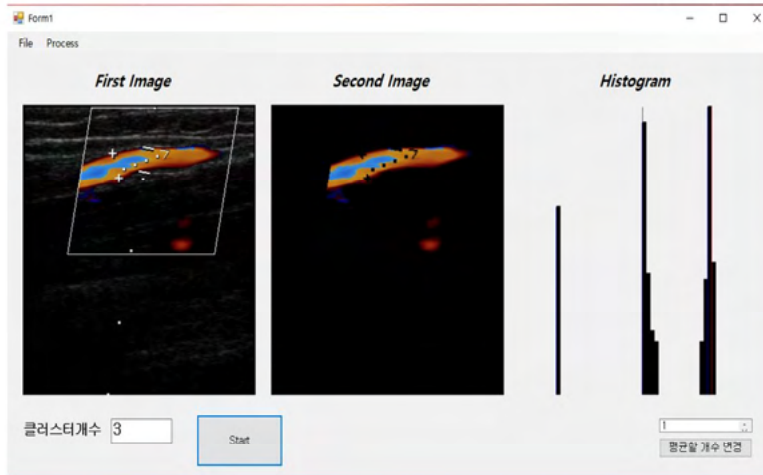
FCM 알고리즘 처리 과정



Fuzzy C-Means

FCM 기반 양자화

신라대학교



FCM 기반 한방 자가 진단 시스템

신라대학교



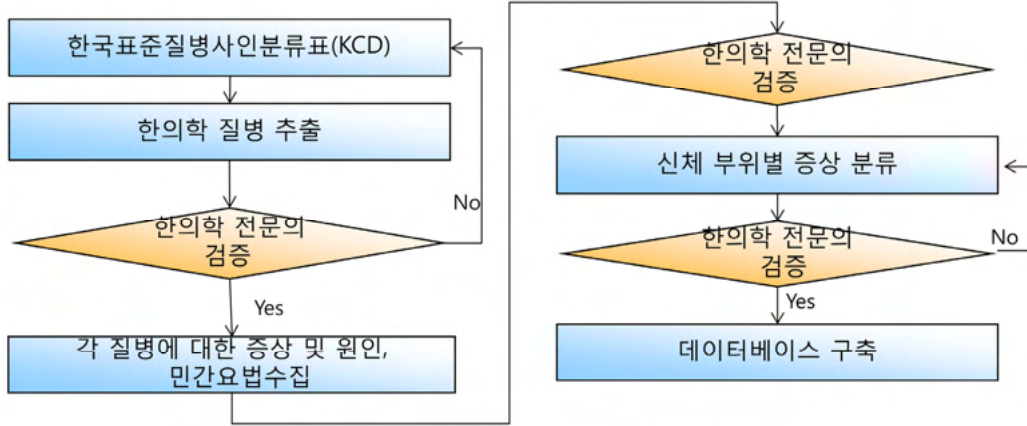
- ✓ 한의학은 수천 년에 걸쳐 인류의 경험과 지혜가 모여져 한국인의 신체적 특성에 맞게 체계화된 동양 의료기술의 결정
- ✓ 한의학을 기준으로 우리 몸에 증상이 발견되었을 때 ,조기에 진단 할 수 있는 지능형 한방 자가 진단 시스템을 제안
- ✓ FCM 알고리즘을 개선하여 질병을 분류하는데 적용하고 퍼지 제어 기법을 이용하여 증상 도출

데이터 베이스 생성 및 구성

신라대학교



◆ 질병 - 증상 및 데이터베이스 구축 과정



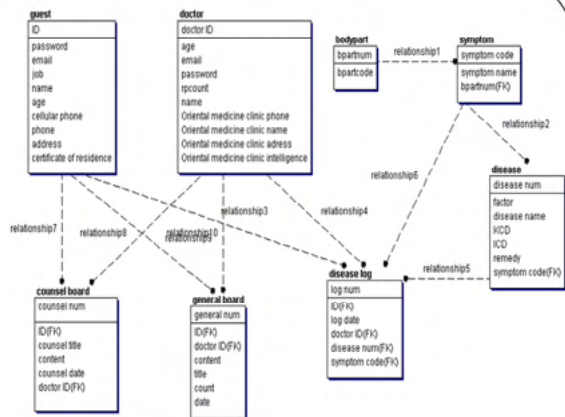
[자료수집 및 데이터베이스 구축과정]

데이터 베이스 생성 및 구성

신라대학교



Entity 명	설명
disease	질병명과 원인 식이요법
symptom	세부 증상
bodypart	부위별 분류(이름)
member	사용자 정보
doctor	전문의 정보
diagnose	진단 결과 누적
Diseaselog	질병도출 기록 저장
counselboard	상담게시판
gernerlboard	일반게시판



[Entity 정의서]

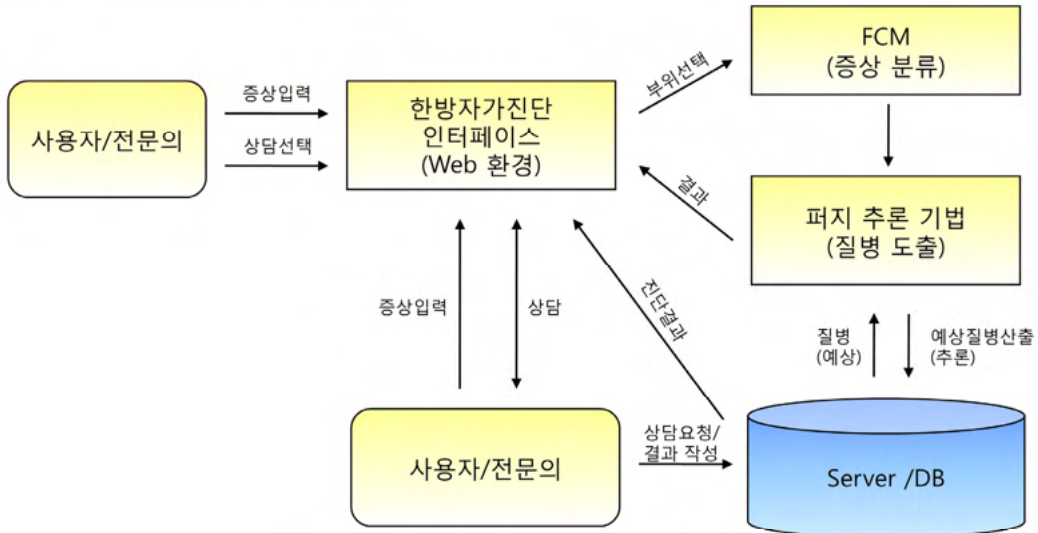
[ER-Diagram]

제안된 한방 자가 진단 시스템

신라대학교



◆시스템의 전체 구성도

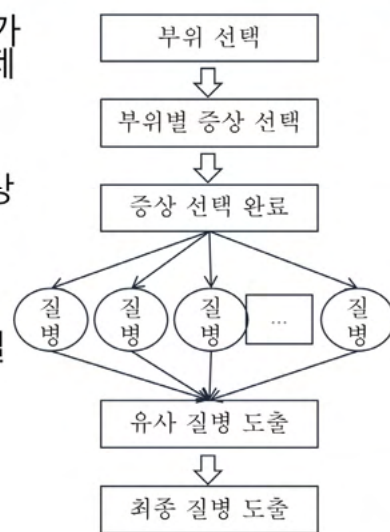


제안된 한방 자가 진단 시스템

신라대학교

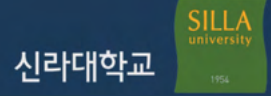


- ◆신체 부위를 자신, 머리, 배, 다리 등 17가지 신체 부위를 성별에 따라 사용자에게 제시
- ◆신체 부위 중에서 사용자가 해당하는 증상 선택
- ◆사용자가 선택한 증상에 대해서 제안된 FCM 알고리즘을 이용하여 유사성 높은 질병 도출



[제안된 한방 자가 진단 처리 과정]

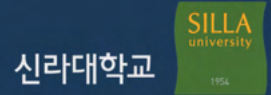
퍼지 기법을 이용한 개선된 FCM 알고리즘



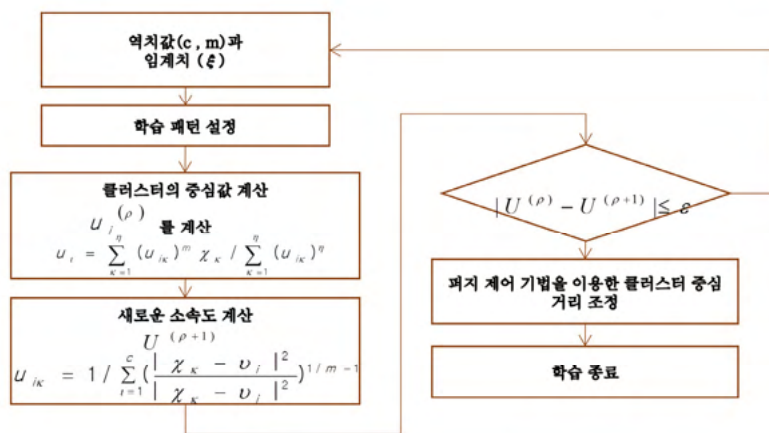
◆ 제안된 한방 자가 진단 시스템의 FCM 알고리즘 개선

1. 사용자가 선택한 증상 패턴과 유사도 측정
2. FCM 알고리즘의 입력 벡터와 군집 중심과의 거리를 계산
3. 클러스터의 중심 거리와 사용자가 입력 벡터의 소속도를 계산하여 질병 패턴 중심과 거리 값 계산

퍼지 기법을 이용한 개선된 FCM 알고리즘



◆ 퍼지 추론 기법을 이용한 개선된 FCM 알고리즘 과정



[개선된 FCM 알고리즘 순서도]

퍼지 기법을 이용한 개선된 FCM 알고리즘

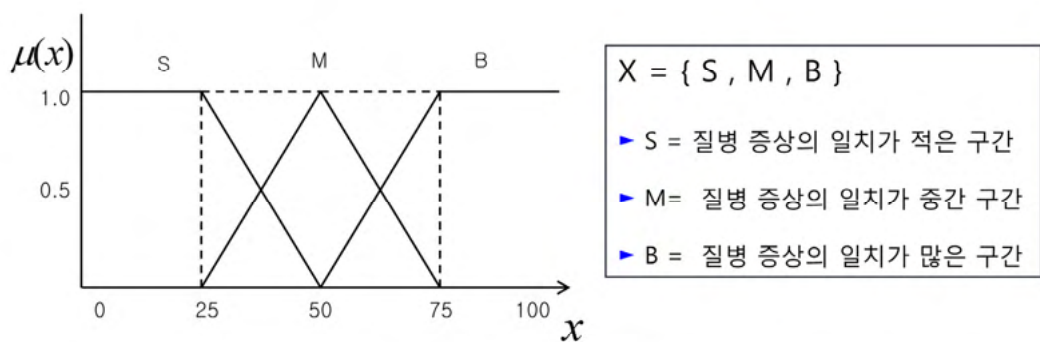
◆ 퍼지 제어 기법을 이용한 클러스터 중심 거리(D) 조정

$$D_k = \sum_{i=1}^c (|x_k - v_i|^2) \times (1.0 - Ot)$$

- ▶ X_k = 사용자가 선택한 증상 벡터
- ▶ V = 대표 증상 중심 벡터
- ▶ Ot = 증상들과의 거리 소속도 평균 벡터 값

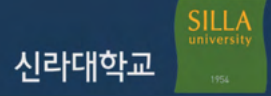
퍼지 기법을 이용한 개선된 FCM 알고리즘

◆ 사용자가 선택한 증상과 질병의 증상 일치 소속함수

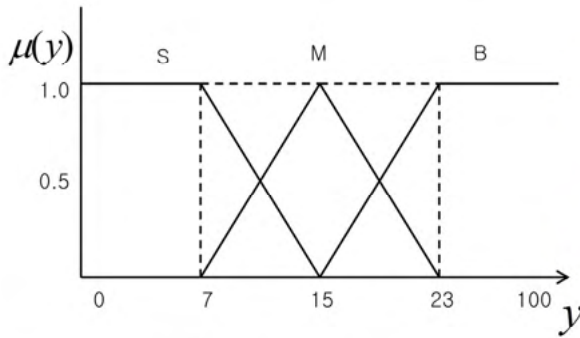


[사용자가 선택한 증상과 질병의 증상 일치 소속함수]

퍼지 기법을 이용한 개선된 FCM 알고리즘



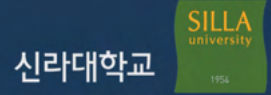
◆사용자가 선택한 증상 수에 대한 소속함수



- $X = \{ S, M, B \}$
- ▶ S = 선택한 증상의 수가 적은 구간
 - ▶ M = 선택한 증상의 수가 중간 구간
 - ▶ B = 선택한 증상의 수가 많은 구간

[사용자가 선택한 증상 수에 대한 소속 함수]

퍼지 기법을 이용한 개선된 FCM 알고리즘



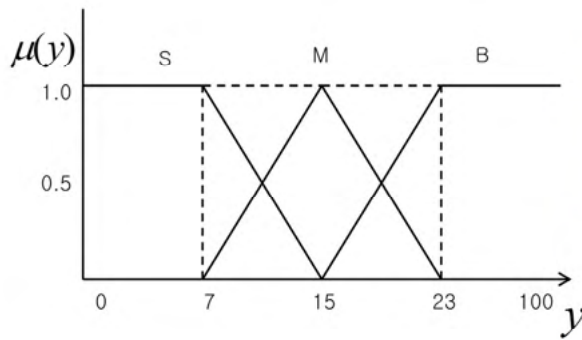
◆제안한 퍼지 추론 규칙

<p>$X = \{ S, M, B \}$</p> <p>S = 증상의 일치 적음 M = 증상의 일치 보통 B = 증상의 일치 많음</p>
<p>$Y = \{ S, M, B \}$</p> <p>S = 선택한 증상 수가 적음 M = 선택한 증상 수가 보통 B = 선택한 증상 수가 많음</p>
<p>$OT = \{ S, M, L \}$</p> <p>S = 사용자가 선택한 증상과 가까운 거리 M = 사용자가 선택한 증상과 보통 거리 L = 사용자가 선택한 증상과 먼 거리</p>

규칙1	If X is S and Y is S Then W is S
규칙2	If X is S and Y is M Then W is S
규칙3	If X is S and Y is B Then W is M
규칙4	If X is M and Y is S Then W is S
규칙5	If X is M and Y is M Then W is M
규칙6	If X is M and Y is B Then W is L
규칙7	If X is B and Y is S Then W is M
규칙8	If X is B and Y is M Then W is L
규칙9	If X is B and Y is B Then W is L

퍼지 기법을 이용한 개선된 FCM 알고리즘

◆OT 값을 계산하는 출력 소속함수



$$X = \{ S, M, L \}$$

- ▶ S = 사용자가 선택한 증상과 거리가 가까운 구간
- ▶ M = 사용자가 선택한 증상과 거리가 보통 구간
- ▶ L = 사용자가 선택한 증상과 거리가 먼 구간

개발 화면



[로그인 화면]



[회원 가입 화면]

개발 결과 분석

신라대학교



[증상 선택 초기화면]



[질병 도출 결과 화면]

연구내용

신라대학교



◆ 지능형 한방 자가진단 웹 서비스 개발



퍼지 이진화

일반 이진화?

- 최적의 임계치를 결정하기 위해 양봉의 히스토그램에서 골짜기를 선택해 적합한 임계치 설정
- 배경과 물체의 밝기 분포가 양봉의 특성을 보이지 않는 경우 임계치 찾기 어려움

퍼지 이진화?

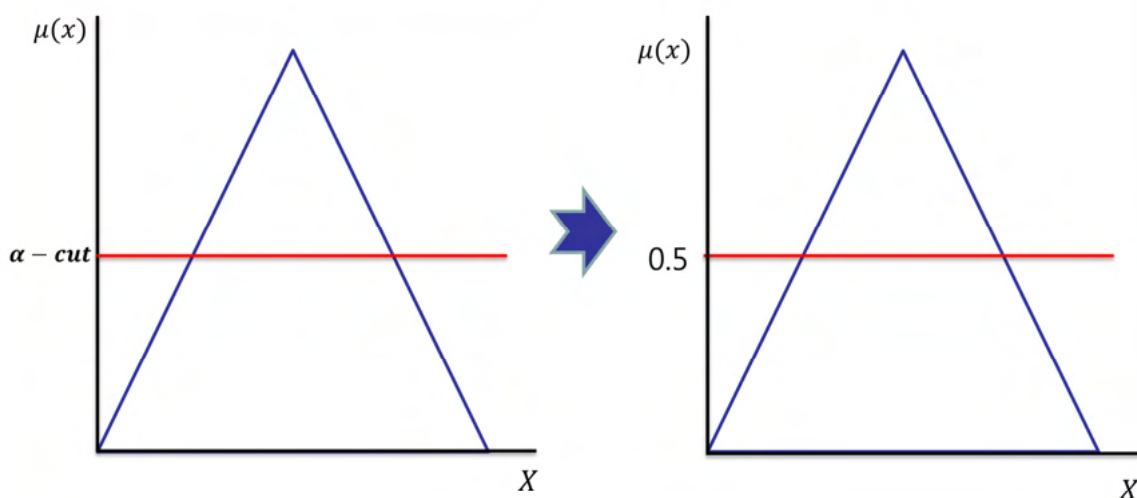
- 삼각형 타입의 소속함수에 적용된 소속도를 이용해 $\alpha - cut$ 을 설정해 이진화

퍼지 이진화 문제점

- 소속도를 경험적인 $\alpha - cut$ 값을 설정해 이진화

FUZZY 이진화 문제점

삼각형 소속함수



FCM 알고리즘을 이용한 퍼지 이진화 개선 방법

신라대학교



FCM 알고리즘

- 단계 2 : 출력 가중치 벡터

$$v_i(t) \leftarrow \frac{\sum_{k=1}^N u_{ik}^m(t)x_k}{\sum_{k=1}^N u_{ik}^m(t)}$$

- 단계 3 : 클러스터 소속도 계산

$$u_{ik}(t+1) = \frac{1}{\sum_{j=1}^c \frac{\|x_k - v_i(t)\|^{2/(m-1)}}{\|x_k - v_j(t)\|}}$$

- 단계 1 : 각 인수들을 초기화한다.
- 단계 2 : 클러스터의 중심값을 계산한다. ($v_i(t)$ 는 클러스터의 중심값, x_k 는 입력 벡터의 값, N 은 영상 전체의 픽셀수, i 는 클러스터의 수, m 은 지수 가중치이다.)
- 단계 3 : 클러스터의 소속도를 계산한다. (c 는 클러스터의 갯수이다.)

FCM 알고리즘을 이용한 퍼지 이진화 개선 방법

신라대학교



FCM 알고리즘

- 단계 4 : 오차 한계치 계산

$$\|U(t+1) - U(t)\| \leq \epsilon$$

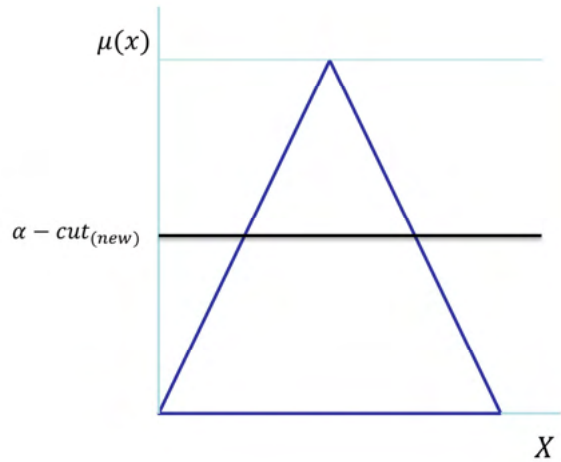
- 단계 4 : 오차 한계치를 계산한다.
- 단계 5 : 현재 소속도 ($U(t+1)$)와 이전 소속도 ($U(t)$)의 차이가 오차 한계치 (ϵ)보다 작거나 같으면 알고리즘을 종료하고 그렇지 않으면 단계2로 가서 반복 수행한다.

FCM 알고리즘을 이용한 퍼지 이진화 개선 방법

개선된 $\alpha - cut$ 값

$$\alpha - cut_{(new)} = \frac{C_{avg}}{255}$$

FCM을 이용해 구해진 각 클러스터링들의 중심 값의 평균을 구한 후 위의 식과 같이 새로운 $\alpha - cut_{(new)}$ 을 구한다. (C_{avg} 는 FCM 알고리즘을 적용하여 클러스터링 된 값들의 평균값이다.)



실험 및 결과 분석



(a) 원 영상

(b) 오토이진화

(c) 퍼지이진화

(d) FCM 기반 퍼지 이진화

4. Deep FCM

Deep FCM : 2 layer FCM

Procedure FCM_2LayerClustering (x) **returns** prototypes and partition matrix

Input data $x = \{x_1, x_2, \dots, x_k\}$

Local fuzzification parameter : m

threshold : ε

norm : $\|\cdot\|$

INITIALIZE-PARTITION-MATRIX

t \leftarrow 0

repeat

for i=1:c **do**

$$v_i(t) \leftarrow \frac{\sum_{k=1}^N u_{ik}^m(t) x_k}{\sum_{k=1}^N u_{ik}^m(t)} \text{ compute prototypes}$$

Deep FCM : 2 layer FCM

신라대학교



for i=1:c **do**

for k=1:N **do**

$$u_{ik}(t+1) = \frac{1}{\sum_{j=1}^c \frac{\|x_k - v_j(t)\|^{2/(m-1)}}{\|x_k - v_i(t)\|^{2/(m-1)}}}$$

update partition matrix

t ← t + 1

until $\|U(t+1) - U(t)\| \leq \varepsilon$

repeat

for s=1:cc **do**

$$w_i(t) \leftarrow \frac{\sum_{i=1}^c q_{si}^m(t) u_i}{\sum_{i=1}^c q_{si}^m(t)}$$

2 Layer compute prototypes

Deep FCM : 2 layer FCM

신라대학교



for s=1:cc **do**

for i=1:c **do**

$$q_{si}(g+1) = \frac{1}{\sum_{r=1}^{cc} \frac{\|q_i - w_s(g)\|^{2/(m-1)}}{\|q_i - w_r(g)\|^{2/(m-1)}}}$$

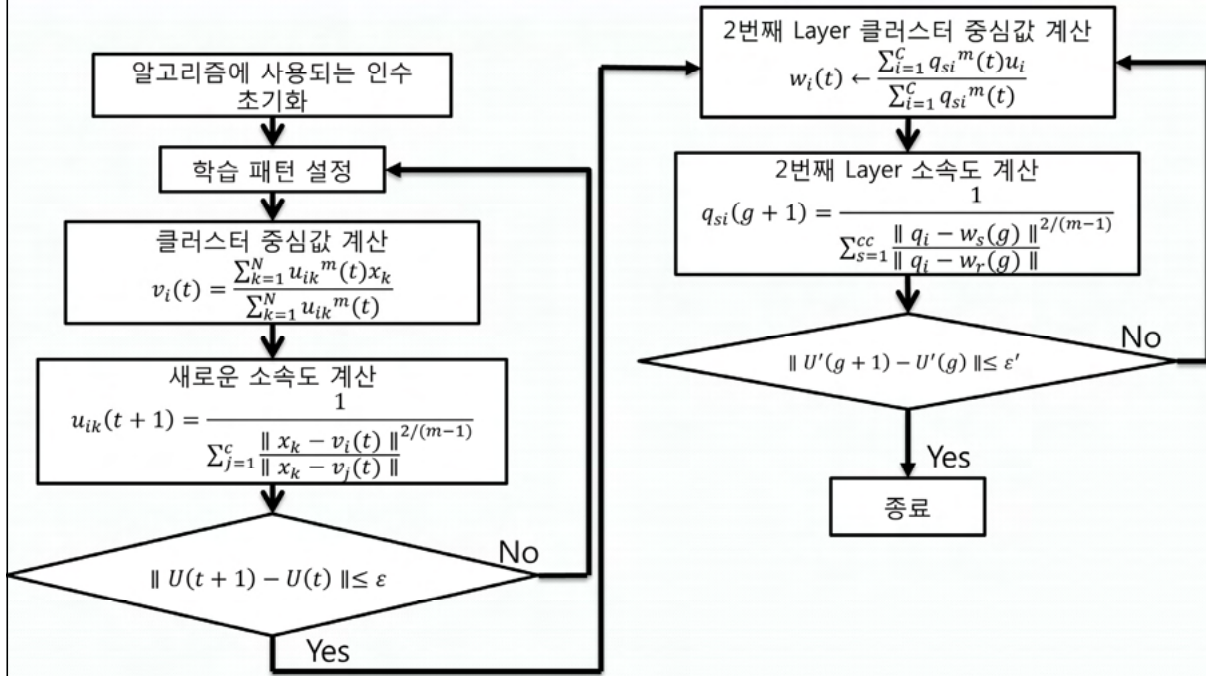
update 2 Layer partition matrix

g ← g + 1

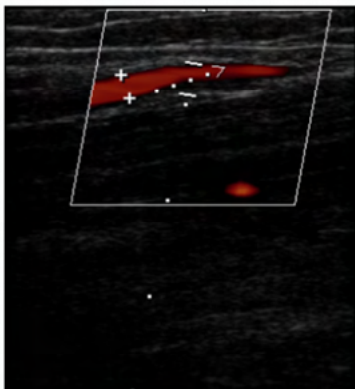
until $\|U'(g+1) - U'(g)\| \leq \varepsilon'$

return w, q

Deep FCM : 2 layer FCM



상완동맥 추출 결과



원본 상완동맥 초음파 영상



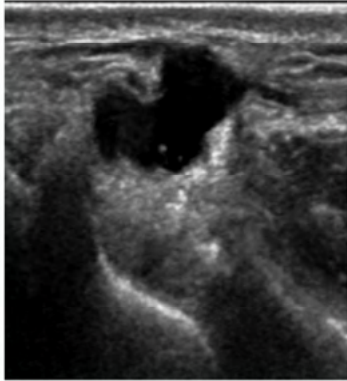
FCM 처리 결과



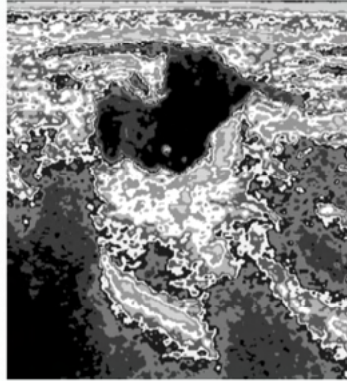
2 layer FCM 처리 결과

결절종 추출 결과

신라대학교



원본 결절종 초음파 사진



FCM 처리 결과



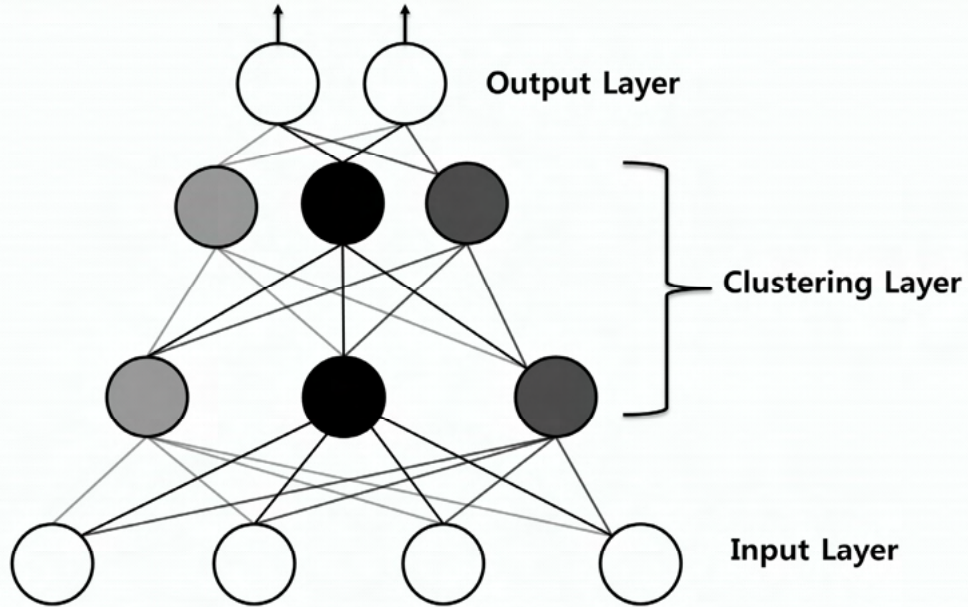
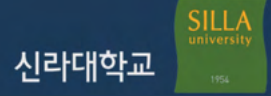
2 Layer FCM 처리 결과

신라대학교

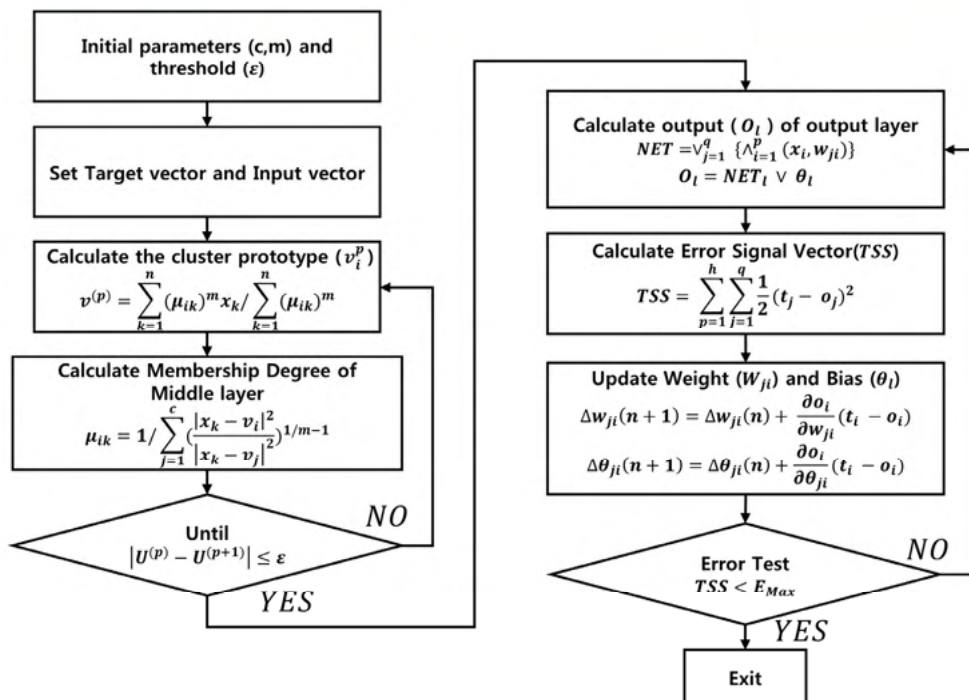
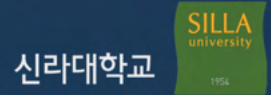


5. Deep Fuzzy Learning Algorithm

Deep Fuzzy Learning Algorithm

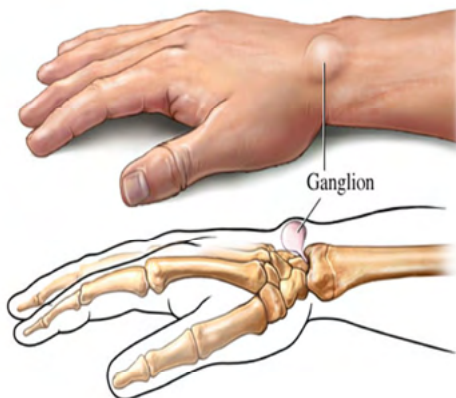


Deep Fuzzy Learning Algorithm



6. 개선된 퍼지 스트레칭 기법과 퍼지 클러스터링 기법을 이용한 초음파 영상에서의 결절종 추출

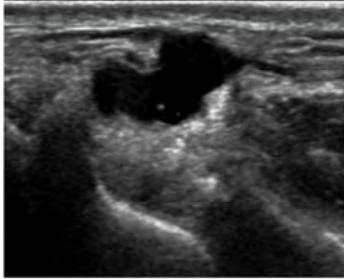
▶ 결절종이란?



- 결절종은 섬유성 피막 내에 액체가 담긴 낭포성 종양으로 주로 손목 또는 발목에 발생합니다.
- 결절종은 업무 및 운동 등과 같이 일상 생활을 하면서 관절부에 발생하며, 결절종이 발생하는 원인은 아직까지 정확히 알려지지 않았습니다.

▶ 기존 연구의 문제점

신라대학교



<기본 스트레칭>



<삼각형 타입 퍼지 스트레칭>

기본 스트레칭의 단점

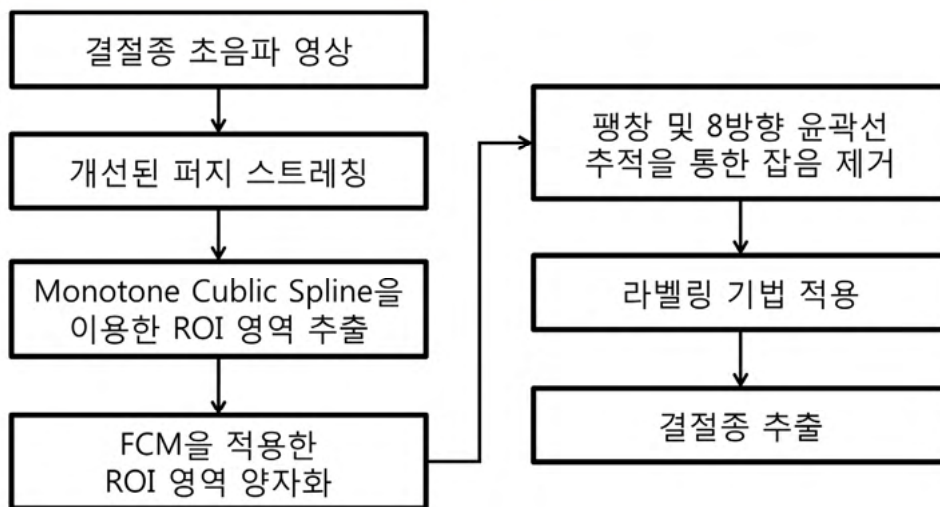
영상 전체의 평균 명암도의 값이 낮은 경우 스트레칭 적용시 불필요한 영역까지 명암대비가 강조되고 결절종 영역의 명암 대비가 적게 반영되는 문제가 발생한다

삼각형 타입 퍼지 스트레칭의 단점

$\alpha = Cut$ 을 경험적으로 설정하므로 명암 대비가 낮게 반영된다. 따라서 다른 영역의 명암 대비가 높게 나타나는 경우가 발생한다.

▶ 결절종 영역 추출 순서도

신라대학교



▶ 개선된 퍼지 스트레칭 순서

신라대학교



영상의 평균 밝기 값

$$X_m = \sum_{i=0}^{M \times N} \frac{X_i}{M \times N}$$

최대 거리 값, 최소 거리 값

$$D_{max} = |X_{max} - X_m|$$

$$D_{min} = |X_m - X_{min}|$$

밝기 값 조정률(δ)

if ($X_m > 128$) then $\delta = 255 - X_m$
 else if ($X_m \leq D_{min}$) then $\delta = D_{min}$
 else if ($X_m \geq D_{max}$) then $\delta = D_{max}$
 else $\delta = X_m$

밝기 값의 최대,최소를 계산

$$I_{max} = X_m + \delta$$

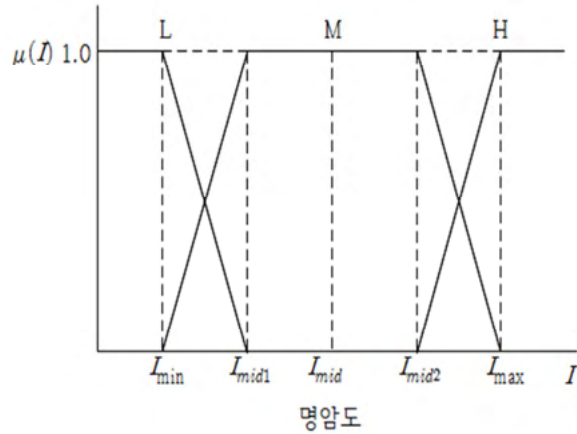
$$I_{min} = X_m - \delta$$

중간 밝기 값 및 새로운 중간 밝기 값 설정

$$I_{mid} = \frac{I_{max} + I_{min}}{2}$$

$$I_{mid1} = \frac{I_{mid} + I_{min}}{2}$$

$$I_{mid2} = \frac{I_{mid} + I_{max}}{2}$$



▶ 팽창

신라대학교

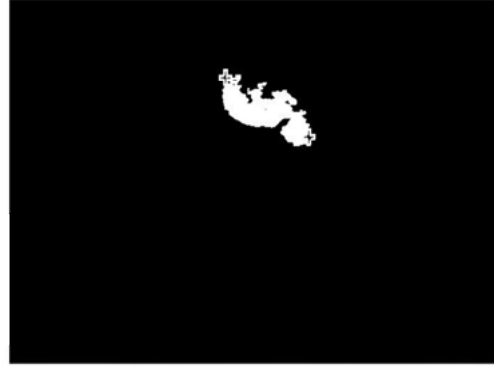
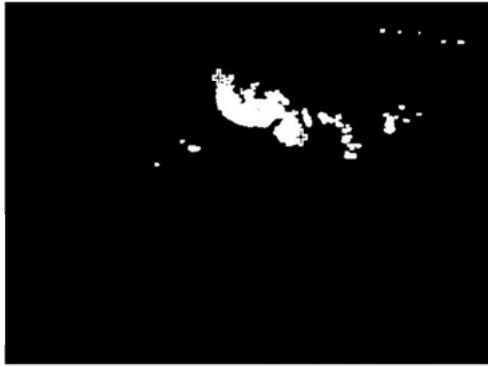


이진화 영상

팽창 기법 적용 결과

▶ 8방향 윤곽선 추적기법

신라대학교

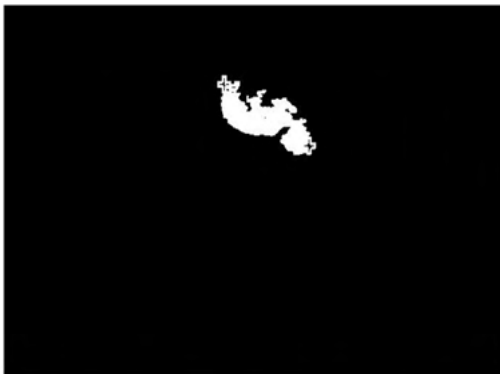


팽창 기법이 적용된 영상

8방향 윤곽선 추적 적용 영상

▶ Labeling

신라대학교

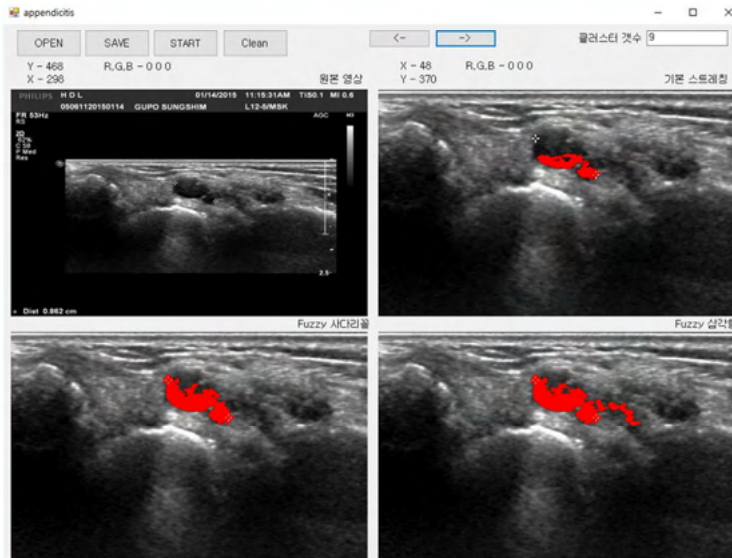


8방향 윤곽선 추적 적용 영상

Labeling 적용 후 결절종 영역 추출

▶ 실험 영상 & 환경

신라대학교



실험 영상

- 프로세서 : Intel(R) Core(TM) i5 CPU @ 2.80GHz
- 메모리(RAM) : 8.00GB
- 시스템 : 64Bit Window10
- Visual Studio 2013 C#

실험 환경

▶ 실험 결과 분석

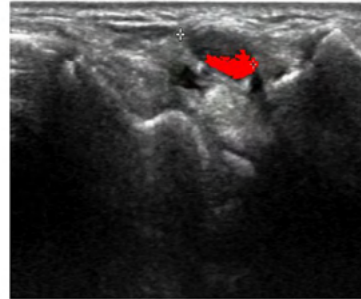
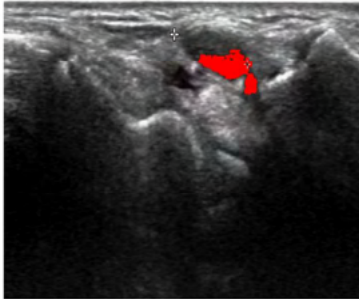
신라대학교



기본 스트레칭과 FCM 양자화 방법	$\text{성공률} = \frac{11}{11 + 9} \times 100 = 55(\%)$
삼각형 타입의 퍼지 스트레칭과 FCM 양자화 방법	$\text{성공률} = \frac{15}{15 + 5} \times 100 = 75(\%)$
사다리꼴 타입의 퍼지 스트레칭과 FCM 양자화 방법	$\text{성공률} = \frac{17}{17 + 3} \times 100 = 85(\%)$

▶ 실험 결과 분석

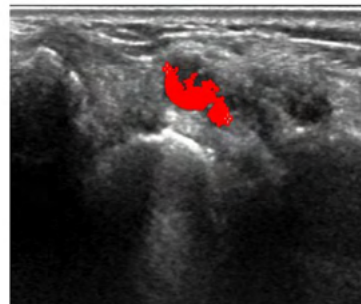
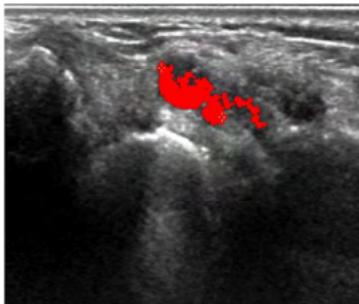
신라대학교



기존의 방법에서 실패영상



제안된 방법



**유전 알고리즘 :
좋은 것이 남는 진화의 법칙**

13:00~14:50

우영운 교수(동의대학교)

유전 알고리즘: 좋은 것이 남는 진화의 법칙



출처: <http://www.indaily.co.kr/client/news/newsView.asp?nBcate=F1009&nMcate=M1008&nIdx=20506&cpage=1&nType=1>

1

유전 알고리즘: 좋은 것이 남는 진화의 법칙

- 강의 자료 출처



2

유전 알고리즘: 좋은 것이 남는 진화의 법칙

• 학습 목표

- 유전 알고리즘이란?
- 유전 알고리즘의 구체적 예
- 유전 알고리즘의 응용 분야
- 프로그래밍 언어를 이용한 유전 알고리즘 시연

3

1. 유전 알고리즘이란?

- 유전 알고리즘이 생물의 유전과 진화에서 학습한다는 것은, 대상 문제의 해를 세대교체를 반복함에 따라 점점 좋은 것으로 변화시켜 나간다는 것을 의미.
- 이 개념은 1960년대부터 있었지만, **홀랜드(John Holland, 1975)**에 의해 개념이 확립되었음.
- 현실에서는 세대교체에 몇 년씩 걸리지만 컴퓨터상에서라면 몇만 세대의 세대교체를 순식간에 실행할 수 있음.
- 이것으로 문제가 해결된다면 고맙겠지만, 초기에는 이론적인 증명이 부족하였고 그 이후에도 논리적인 검증과 확장 연구가 계속 이루어지고 있음.

4

1. 유전 알고리즘이란?



John H. Holland

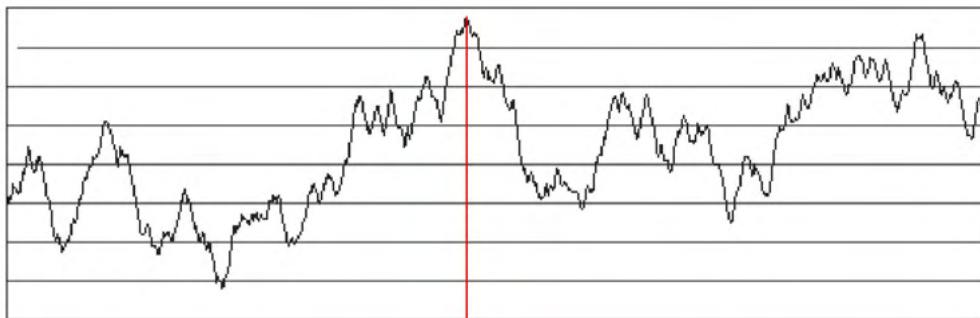
Born in Indiana 1929.
Author of “Adaptation in Natural and Artificial Systems”
written in 1975, providing the basis of genetic algorithms.
Recipient of the McArthur Fellowship.

John Henry Holland (February 2, 1929 – August 9, 2015)

출처: <http://slideplayer.com/slide/8342885/>

5

1. 유전 알고리즘이란?



6

1.1 유전 알고리즘의 개념

- 우선 대상 문제를 세대 교체할 수 있는 모델로 표현할 필요가 있음. 즉, 대상 문제의 특징을 추출하여 몇 가지 **기호열**로 표현하고 이 기호열을 세대교체의 대상으로 함.
- 이 기호열이 유전자에 해당하며 이와 같은 모델화가 바로 유전자 설계에 해당함.
- 유전 알고리즘에서는 수식을 사용하지 않는 대신 코딩에 신경 씀.
- 유전자 설계 형태는 다양한 기법에 따라 개발되어 있으며, 그 설계 형태에 따라서 유전 알고리즘이 적용 여부가 결정됨.
- 유전자 설계와 동시에 해로서의 가치를 평가하는 지표도 필요하며, 이것을 **적응도**라고 함.

7

1.1 유전 알고리즘의 개념

- 세대교체에는 3가지 실행 단계가 있음.
 - ① **선택**: 다양한 유전자를 갖는 개체 중에서 적응도가 높은 것을 선택.
 - ② **교차**: 그 유전자들 간에 유전자 일부를 교환하여 보다 적응도가 높은 유전자를 갖는 개체를 만들어 냄.
 - ③ **돌연변이**: 진화가 정체되지 않도록 가끔 새 유전자를 만들어 끼워 넣음.
- 세부 개념을 정리하면 다음과 같음.
 - **유전자(Gene)**: 대상 문제의 특징을 추출하여 세대교체 대상이 되는 기호열
 - **적응도(Fitness)**: 대상 문제에서 요구되는 가치에 얼마나 근접한지를 나타내는 지표
 - **선택(Selection)**: 많은 개체들로부터 적응도가 높은 것을 선택해 내는 방법
 - **교차(Crossover)**: 개체 간에 유전자를 교환하는 방법
 - **돌연변이(Mutation)**: 적응도에 관계없이 임의로 유전자를 끼워 넣는 방법

8

1.2 유전 알고리즘의 처리 순서

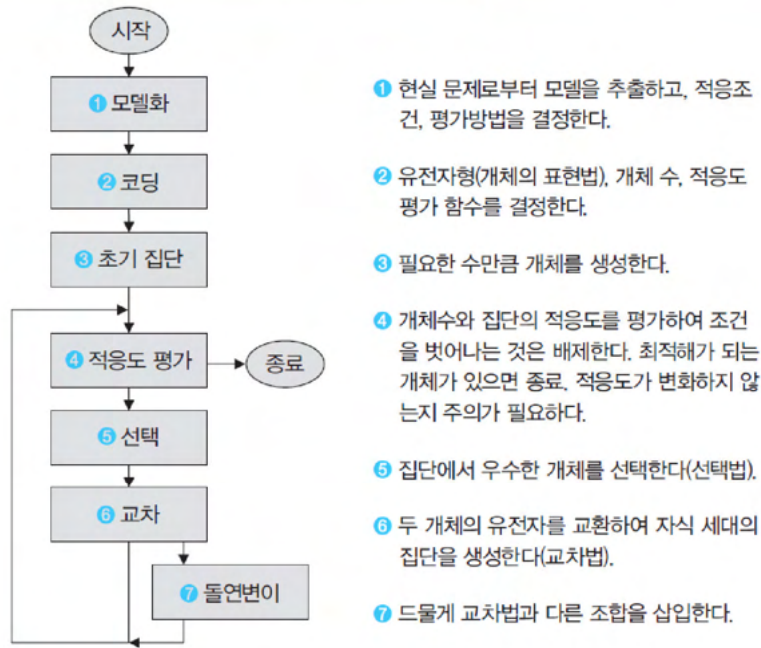


그림 4-1 유전 알고리즘의 처리 순서

1.2 유전 알고리즘의 처리 순서

- 유전 알고리즘을 적용하는 순서
 - ① **모델화:** 대상 문제의 특징을 추출하여 목표 상태를 정의.
 - ② **코딩:** 유전자와 적응도 평가 함수를 정의.
 - ③ **초기 집단:** 적당한 유전자를 갖는 개체를 필요한 개수만큼 생성.
 - ④ **적응도 평가:** 적응도를 평가. 목표 상태가 되면 종료.
 - ⑤ **선택:** 적응도가 높은 유전자를 갖는 개체를 선택.
 - ⑥ **교차:** 개체 간에 유전자를 교환.
 - ⑦ **돌연변이:** 필요에 따라 적응도에 관계없이 새 유전자를 끼워 넣음.

1.2 유전 알고리즘의 처리 순서

- 이 처리 순서에서 ④~⑦을 반복하며, ④에서 목표 상태가 되면 종료되지만 그렇다는 보장은 없기 때문에 보통 반복하는 횟수를 지정해 둬.
- 일정 횟수를 반복하여도 기대하는 해가 얻어지지 않으면 다시 한번 더 반복하거나 초기 집단을 재설정하여 다시 실행.
- 그래도 해결되지 않는 경우에는 코딩 단계부터 새로 고쳐야 할 가능성도 있음.

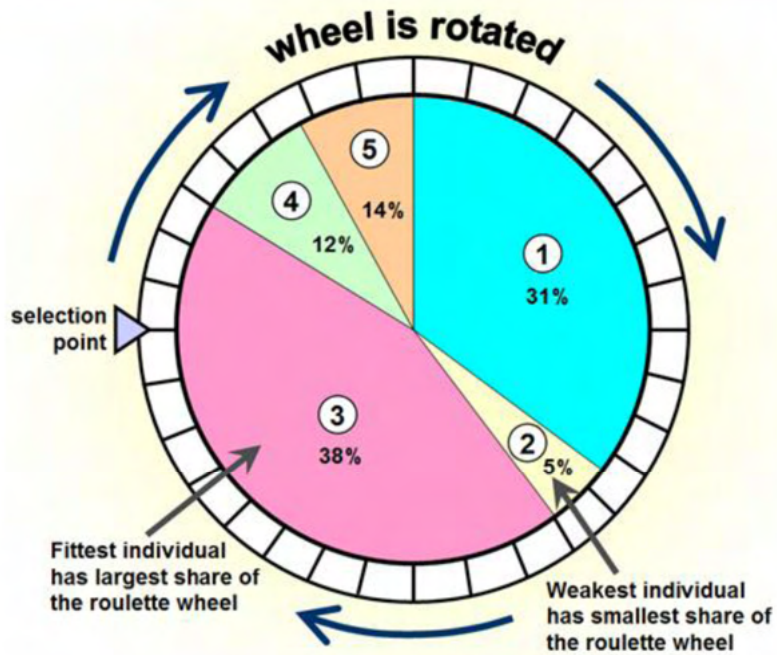
11

1.3 선택법

- 세대교체에서는 적응도가 높은 유전자를 갖는 개체를 선택하는데, 선택하는 방법은 다음과 같이 몇 가지가 있음.
 - ✓ **엘리트 보존:** 적응도가 가장 높은 것을 그대로 다음 세대에 전달하는 것.
 - ✓ **룰렛 선택:** 자식을 적응도에 비례하는 확률로 선택한다(확률은 $p_i = f_i / F$, 단 f_i 는 개체 i 의 적응도 $F = \sum f_i$). 적응도가 낮다고 해서 버리는 것은 아니지만 실제로는 자식도 유한 개수이므로 적응도가 낮은 개체는 무시됨.
 - ✓ **토너먼트 선택:** 무작위로 선택한 개체 중(보통 2개)에서 가장 적응도가 높은 개체를 선택한다. 보통 높은 적응도를 갖는 것만 남기 때문에 수렴은 빠르지만 **국소해**에 빠질 가능성이 높음.
- 다른 선택법도 있으며 조합하여 사용하는 경우도 있음. 홀랜드가 최초로 제안한 방법은 **룰렛 선택**이지만 그 이후 개선을 위한 연구가 이루어지고 있음.

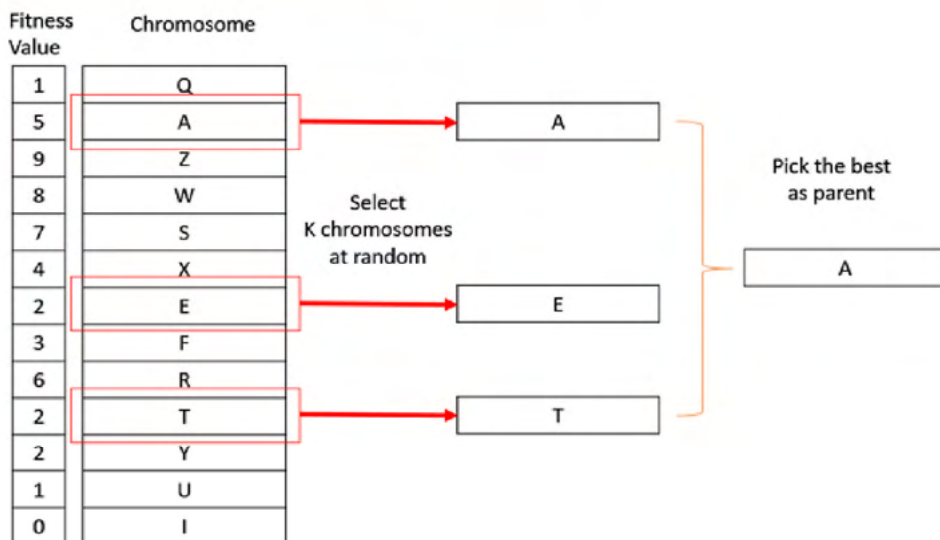
12

1.3.1 룰렛 선택



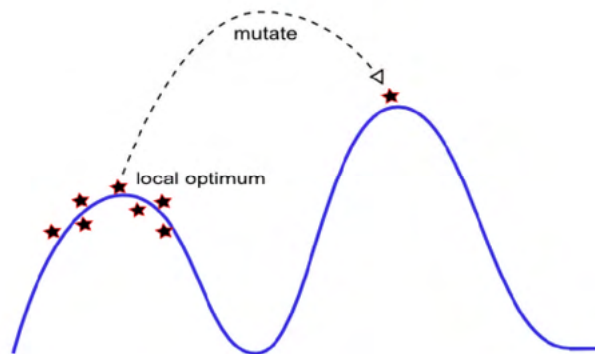
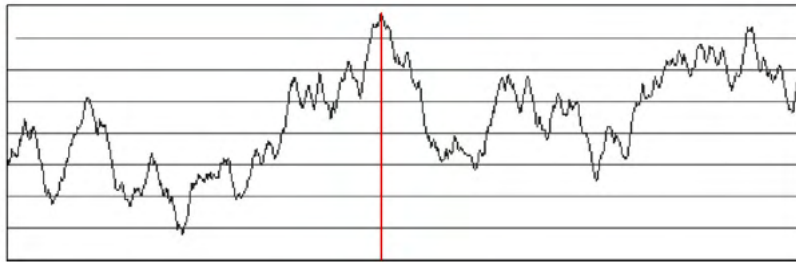
13

1.3.2 토너먼트 선택



14

1.3.3 국소해



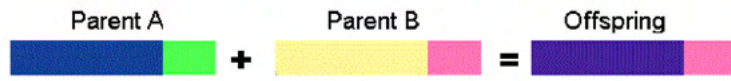
15

1.4 교차법

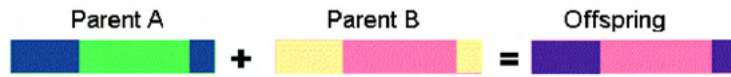
- 선택된 개체의 집단에서 2개씩 조합하여 부모 쌍을 몇 개 만들고, 부모 쌍으로 유전자 교환을 수행.
- 부모 쌍의 가장 일반적인 조합 방법은 적응도가 높은 것과 낮은 것으로 순서대로 해 나가는 것임. 이렇게 함으로써 유전자의 경향이 치우치는 것을 방지할 수 있음.
- 얼핏 보면 **적응도가 높은 것들끼리 조합**하는 편이 효율적인 것 같지만, 이럴 경우 **국소해**로 끝나 버릴 가능성이 높음.
- 오히려 낮은 적응도의 유전자 중에서 보다 좋은 해를 이끌어 낼 수 있는 요소가 숨어 있을 가능성이 있음.
- 다른 조합 방법으로는 무작위로 2개씩 쌍을 만드는 방법도 있음.
- 부모 쌍이 만들어지면 다음의 방법들로 유전자의 일부를 교환.

16

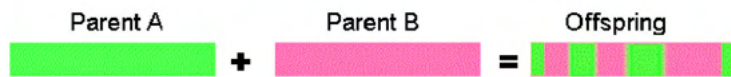
1.4 교차법



단순 교차

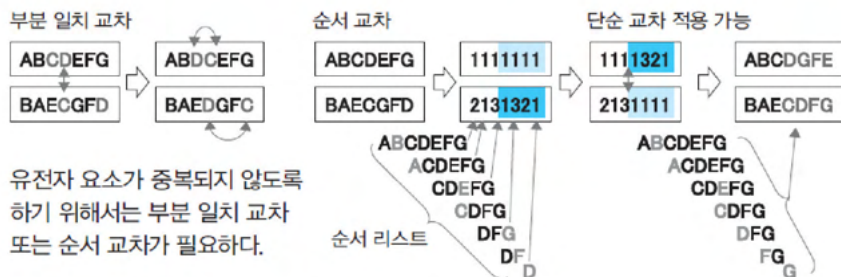
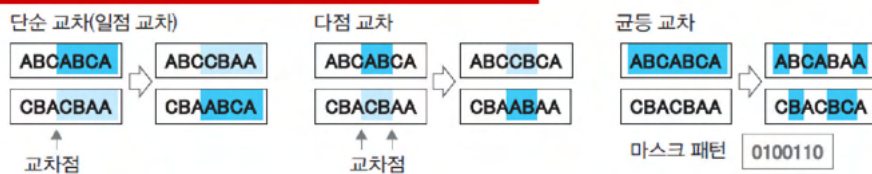


다점 교차



균등 교차

1.4 교차법



서브투어 부분을 상호 간에 그대로 교환(순서 교환)하고, 다시 역순으로 교환(역순 교환)하기 때문에 두 부모로부터 4개의 자식이 생성된다.

그림 4-2 교차법

1.4 교차법

- **일점(single point) 교차(단순 교차):** 적당한 한 곳을 정하여 그 이후의 유전자를 교환.
- **다점(multiple point) 교차:** 부분적으로 여러 개의 유전자를 일제히 교환. 몇 군데라도 상관없음.
- **균등(uniform) 교차:** 부모 쌍으로부터 마스크 패턴에 따라 유전자를 복사. 예를 들어, 0이면 부모 1로부터, 1이면 부모 2로부터 유전자를 복사.
- **부분 일치(partially matched) 교차:** 순서가 중요한 문제가 되는 경우에는 다점 교차로 교차 쌍을 정하고 각 개체 내에서 바꿔 넣음.
- **순서(order) 교차:** 유전자의 중복을 허용하지 않는 경우에는 유전자 자체 교환이 아니고 순서대로 바꾸어 넣어 교차를 수행.
- **서브투어 교환(subtour exchange) 교차:** 복수의 개체에 공통적인 부분이나 좋은 성질을 갖는 부분을 유지하면서 교차 수행. **스키마**라고도 함.

19

2. 유전 알고리즘의 구체적 예

- **재산 분배 문제**
 - ✓ **수식화가 어려운 조합 최적화 문제**로서 다양한 가치를 갖는 물건을 여러 사람에게 정해진 비율로 나누어 주는 재산 분배 문제를 생각.
 - ✓ 아버지의 유언은 재산을 장남, 차남, 삼남에게 **4:2:1의 비율**로 물려준다는 것이었으며, 자산 가치가 다른 7개의 물건이 재산으로 남아 있음.

20

2. 유전 알고리즘의 구체적 예

- 얼핏 보기에는 재산 총액과 각자가 취할 몫에 대한 비율로 간단하게 분배할 수 있을 것으로 여겨짐.
- 그러나 재산 물건이 연속적으로 분할될 수 있는 형태가 아니기 때문에 제대로 분배하려고 하면 두더지잡기 게임처럼 들쭉날쭉한 상태에 빠짐.
- 그래서 생각할 수 있는 분배 몇 가지를 시험 삼아 해 보고, 기대 비율이 가장 근접한 분배에 따르도록 함.
- 여기서 유전 알고리즘의 적용 가치가 드러남.

21

2. 유전 알고리즘의 구체적 예

재산 분배 문제 아버지의 유언은 재산을 장남, 차남, 삼남에게 4:2:1의 비율로 물려준다는 것이었으며, 자산 가치가 다른 7개의 물건이 재산으로 남아 있다.



그림 4-3 재산 분배 문제

22

2.1 재산 분배 문제 해설

- 앞의 그림에서 세대교체의 모습을 자세히 살펴보자.
- ③의 초기 집단은 적당히 선택한 4개의 유전자로 이루어짐.
- 첫 번째 유전자는 물건 1은 A, 물건 2는 B, 물건 3은 C,..., 물건 7은 A가 상속받는 것을 나타냄.
- 두 번째 유전자를 같은 방식으로 보면 A는 물건 1, 3, 7을, B는 물건 2, 4를, C는 물건 5, 6을 상속받는 것을 나타냄.
- 세 번째 유전자는 A가 3, 6, B가 1, 4, 7, C가 2, 5를, 네 번째 유전자는 A가 3, 6, 7, B가 2, 5, C가 1, 4를 상속받는 것을 나타냄.

23

2.1 재산 분배 문제 해설

- ④에서는 이 유전자들의 적응도를 평가하고 있음.
- 각 유전자가 나타내는 분배를 보면, 첫 번째는 A가 물건 합계 12억, B가 7억, C가 9억으로 분배되며, 두 번째는 A가 11억, B가 6억, C가 11억, 세 번째는 A가 9억, B가 12억, C가 7억, 네 번째는 16억, 7억, 5억으로 분배.
- 이 분배들이 목표인 4:2:1이라는 비율에 어느 정도 가까운지가 적응도이지만, 여기서는 엄밀한 계산은 생략하고 적응도가 높은 순으로 순서를 매김.
- 이 경우에는 4번째 유전자가 가장 적응도가 높고 이어서 첫 번째, 두 번째, 세 번째의 순서로 결정.

24

2.1 재산 분배 문제 해설

- ⑤에서는 선택을 수행. 엘리트 보존에 따라 가장 적응도가 낮은 유전자를 가장 높은 것으로 바꿔 놓음.
- 즉, 세 번째 유전자를 네 번째 유전자로 교체.
- 이렇게 하면 보통 적응도가 가장 높은 유전자는 2개가 됨. 그리고 이 상태로 적응도가 가장 높은 것과 가장 낮은 것을 한 쌍으로 정하고, 그 다음으로 높은 것과 낮은 것을 다시 한 쌍으로 정함.
- 이 경우에는 세 번째와 두 번째 유전자가 한 쌍(쌍 1), 네 번째와 첫 번째 유전자가 다른 한 쌍(쌍 2)이 됨.
- 여기서서는 단순 교차를 수행하고 교차점을 쌍 1은 2(뒤로부터 2개 요소), 쌍 2는 4(뒤로부터 4개 요소)로 함.

25

2.1 재산 분배 문제 해설

- ⑥에서는 실제로 단순 교차(일점 교차)를 수행. 즉, 쌍 1은 2개 유전자의 뒷부분 2개 요소를 교환하고, 쌍 2는 4개 요소를 교환.
- ④로 돌아가 유전자 조작에 따라 생성된 4개의 자식 유전자에 대한 적응도를 평가. 그 결과, 적응도가 가장 높은 것은 두 번째 자식 유전자였으며, A, B, C의 분배가 17억, 6억, 5억이 되었음.
- 가장 낮은 것은 세 번째 자식 유전자이므로 선택(⑤)에서 세 번째를 두 번째로 교체하고, 다시 2조의 쌍을 만들어 교차점을 2와 4로 하여 교차(⑥)를 수행.
- 다시 ④로 돌아가 손자 유전자의 적응도를 평가하면 적응도가 가장 높은 것은 16억, 7억, 5억으로 처음 상태로 돌아간 것처럼 보이지만, 적응도가 가장 낮은 것도 14억, 6억, 8억이 되었음.

26

2.1 재산 분배 문제 해설

- 처음 상태와 비교하면 전체적으로는 비교적 적응도가 높은 것들로만 재편되었다는 것을 알 수 있음.
- 더욱이 선택(⑤)을 수행하여 가장 적응도가 낮은 첫 번째 손자 유전자를 네 번째 손자 유전자를 교체하여 쌍을 만들고, 교차(⑥)를 수행하면 가장 적응도가 낮은 첫 번째 증손자 유전자도 14억, 7억, 7억이 되므로 전체적으로도 엘리트들만의 집단이 되었음.
- 가장 적응도가 높은 두 번째 증손자 유전자는 17억, 7억, 4억이므로 목표인 4:2:1의 분배인 16억, 8억, 4억과 비교하면 A가 1억 더 받고, B가 1억 적은 결과가 되었지만 그런대로 분배가 잘 된 것으로 볼 수 있음.

27

2.1 재산 분배 문제 해설

- 이 이후에 몇 번 정도 더 세대교체를 수행하면 목표하는 분배가 될 수도 있겠지만 보장할 수는 없음.
- 관점을 바꾸어 보면, 엘리트들만의 유전자는 더 이상 발전할 수 없는 곳까지 와 버렸기 때문에 세대교체를 수행해도 변화가 없음.
- 그런 상태가 되었을 경우 엘리트라고 할 수 없을 정도로 적응도가 낮은 유전자를 넣어 보면, 일시적으로는 전체 적응도가 낮아 지더라도 머지않아 이전보다 더 좋은 결과가 얻어질 수도 있음.
- 이 부분에 대한 이론적 근거가 부족한 것이 아쉽지만 적용해 보면 확실히 좋아지기도 함.
- 적응도 계산이 필요하지만 세대교체 자체는 동일한 순서로 반복 되므로 매우 편리한 방법임.

28

3. 유전 알고리즘의 응용 분야

- 유전 알고리즘은 이론적 배경이 아직도 완벽하지 않지만 응용 범위는 넓음. 주요 분야는 다음과 같음.
 - ✓ **조합 최적화 문제:** 주어진 제약 조건 내에서 가장 효과적인 조합을 구함. 배낭 문제(Knapsack Problem; KP)와 순회 외판원 문제 (Traveling Salesperson Problem; TSP)가 유명함.
 - ✓ **배치 설계 문제:** 주어진 기능 블록을 작은 공간에 가장 효과적으로 배치한다. LSI 설계, 상업시설 설계 등.
- 이 외에 **태스크 스케줄링**(많은 공정을 최적 순서로 진행), **제어 문제**(에어컨 온도 제어 등), **계획 문제**(버스 운행 편성, 근무 편성 등) 등의 응용 분야가 있음.

29

3.1 조합 최적화 문제

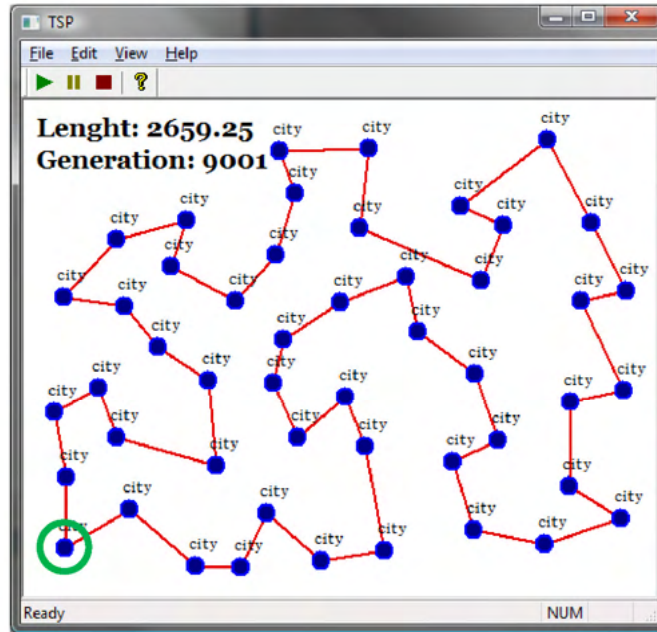
- 배낭 문제(Knapsack Problem : KP)



30

3.1 조합 최적화 문제

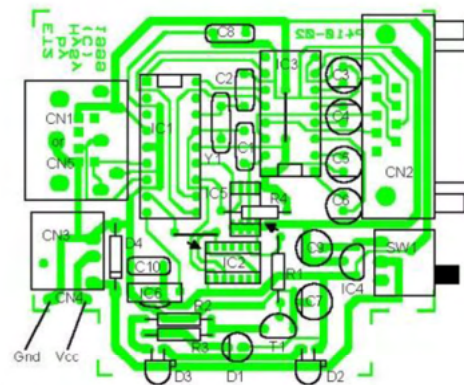
- 순회 외판원 문제(Traveling Salesperson Problem : TSP)



31

3.1 조합 최적화 문제

- PCB 레이아웃 문제



32

3.1 조합 최적화 문제

- **조합 최적화 문제**는 문제의 조합에 해당하는 수치 계산을 끝없이 반복하는 것보다, 유전 알고리즘을 활용하면 더 효율적으로 좋은 해를 구하는 방법이 가능함.
- 여기에서는 2개의 유명한 문제를 살펴봄.
 - ✓ **냅색 문제:** 여러 형태, 무게의 화물들을 배낭(냅색)에 최대한 잘 집어넣는 것. '잘'이라는 의미는 제한 조건에 따라 배낭에 들어간 각 화물이 갖는 가치의 합계가 최대가 되도록 하는 것.
 - ✓ **순회 외판원 문제:** 여러 도시를 중복 없이 잘 돌아다니는 것. '잘'이라는 의미는 각 도시 간의 방문 경로에 대한 비용의 전체 합이 최소가 되도록 하는 것.

33

3.1 조합 최적화 문제

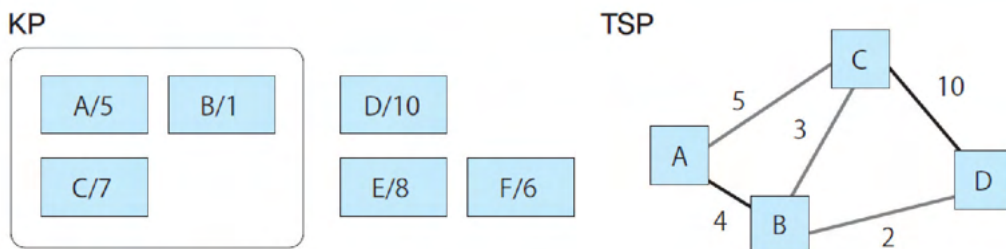


그림 4-4 냇색 문제(KP)와 순회 외판원 문제(TSP)의 개요

- 두 문제 모두 간단한 것처럼 보이지만, 가능한 방법을 모두 하나 하나 조사하려고 할 경우 **화물의 개수나 도시 수가 50개 정도가 되면 경우의 수가 1조가 넘어** 보통의 컴퓨터에서는 도저히 대응할 수 없음.
- 유전 알고리즘을 적용하면, 반드시 최적해를 구한다는 보장은 할 수 없지만, **일정 시간 내에 좋은 해(근사해)를 얻을 수는 있음.**

34

3.1 조합 최적화 문제

- 이 문제들은 보통 다음과 같이 처리.

- ❖ **유전자 결정**

- ✓ **KP:** 각 비트를 각 화물에 대응(1은 있음, 0은 없음)시킨 화물 개수만큼의 길이를 갖는 비트 열. 예를 들어, 1101100010

	1	2	3	4	5	6	7	8	9	10	화물 번호
	1	1	0	1	1	0	0	0	1	0	

- ✓ **TSP:** 도시를 방문하는 순서대로 도시 이름을 나열한 기호 벡터.(단, 길이 없는 도시는 이웃하게 두지 않는다.) 예를 들어, ACBD, 또는 도시 이름에 대하여 그 도시를 방문하는 순서 차레를 나열한 수치 벡터. 제약 사항은 동일함(예를 들어, 1324).

	A	B	C	D	E	F	G	H	I	J	도시 이름
	1	4	10	5	9	6	3	2	8	7	

3.1 조합 최적화 문제

- ❖ **적응도 평가**

- ✓ **KP:** 제약조건 내 개체는 화물 가치의 합계. 조건 외 개체는 가치를 0으로 평가
- ✓ **TSP:** 도시 간 이동 거리의 총합계

- ❖ **초기 집단을 적당히 선택, 선택, 교차를 반복, 가치가 일정 수준 이상이 되면 종료**

- ✓ **KP:** 일반적인 선택, 교차로 충분
- ✓ **TSP:** 모든 도시를 돌아다님 = **유전자의 중복이 없음** → 순서 교차 등

3.2 유전 알고리즘의 사례 및 활용 분야

- 유전 알고리즘
 - <https://www.youtube.com/watch?v=94p5NUogCIM>
 - https://www.youtube.com/watch?v=Yr_nRnqeDp0&t=1s
- NEAT(Neuro Evolution of Augmenting Topologies) 알고리즘
 - https://www.youtube.com/watch?v=C_VUtPNFqYw
 - <https://www.youtube.com/watch?v=qv6UVOQ0F44>
 - <https://www.youtube.com/watch?v=8V2sX9BhAW8>

37

4. 프로그래밍 언어를 이용한 유전 알고리즘 시연

- 실습 목표
 - C 프로그램 언어로 작성한ナップ색 문제 시연
 - 물건의 무게와 가치(100개) 쌍에 대한 데이터 파일
--> knapsack100.txt
 - 넣을 수 있는 최대 무게(30000, 40000, 50000, 70000, 90000)별로 가방에 담기는 물건의 가치를 가장 높이는 조합을 찾는 것이 목적
 - 구현된 프로그램의 수행 결과를 DP(Dynamic Programming) 기법 결과와 비교

38

4. 프로그래밍 언어를 이용한 유전 알고리즘 시연

물건 개수	100	100	100	100	100
최대 무게	30,000	40,000	50,000	70,000	90,000
WOO	29,987	39,994	49,971	69,948	89,662
	960,172	1,103,678	1,220,360	1,382,451	1,464,837
A조	678,496	842,681	956,476	1,291,044	1,420,593
B조	731,566	823,090	894,831	1,099,531	1,068,231
C조	932,861	1,075,527	1,214,552	1,379,127	1,464,837
D조	950,902	1,085,510	1,231,235	1,375,623	1,461,869
E조	933,441	1,102,030	1,223,236	1,386,050	1,459,725
F조	741,975	937,241	1,014,216	1,210,235	1,441,469
DP	962,421	1,115,308	1,232,749	1,395,452	1,464,837

4차 산업 관련 인공지능 기반 기술 워크숍

의사결정트리를 활용한 분류기법

15:00~16:50

김희철 교수(인제대학교)

의사결정트리를 활용한 분류 기법

2018. 08. 24.

김희철

인제대학교 컴퓨터공학부

의사결정트리

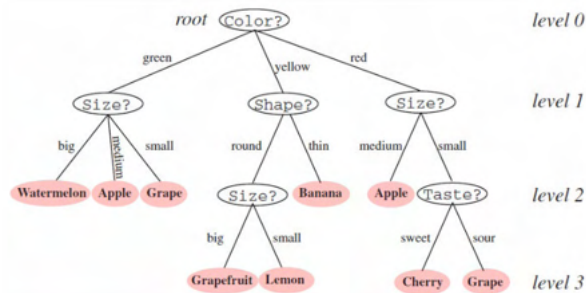
퍼지의사결정트리

의사결정트리 (Decision Tree)

- 정의

의사결정규칙(decision rule)을 트리구조로 도표화하여 관심 대상이 되는 집단을 몇 개의 소집단으로 분류(classification)하거나 예측(prediction)을 수행하는 분석방법

스무고개와 개념이 유사 : 일련의 질문에 근거해서 데이터를 분류해주는 알고리즘



의사결정트리 (Decision Tree)

- 기계학습에서 대표적인 Supervised Learning 방법
- 목표 변수가 범주형인 경우 분류 트리, 수치형인 경우 회귀 트리
- 의사결정트리를 활용하는 분야
 - 분류(Classification)
 - 예측(Prediction)
 - 세분화(Segmentation)
 - 차원 축소, 변수선택, 연속 변수 이산화
- 응용 영역
 - 시장조사, 광고조사, 의학연구, 품질관리 등의 다양한 분야에서 활용
 - 특히 마케팅의 경우, 고객 타겟팅, 고객들의 신용점수화, 캠페인 반응분석, 고객행동예측, 고객 세분화 등에 사용

의사결정트리 (Decision Tree)

- 장점
 - Human readable한 트리구조로 표현되므로 결과를 쉽게 이해할 수 있음 (if..then.. 방식으로 결과 표현)
 - 그 구조가 단순하여 해석이 용이
- 단점
 - 분류 기준값의 경계선 근방의 자료 값에 대해서는 오차가 클 수 있으며(비연속성),
 - 로지스틱 회귀와 같이 각 예측변수의 효과를 파악 어려움
 - Hill climbing 방식 및 Greedy 방식을 사용하는데, 일반적으로 Greedy 방식의 알고리즘은 local optimization에 빠지기 쉬움

학습 데이터 사례

Day	Outlook	온도	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

의사결정트리 알고리즘

- 대표적 알고리즘
 - CHAID(Kass, 1980)
 - CART(Breiman 등, 1984)
 - ID3(Quinlan, 1986)
 - C4.5(Quinlan, 1993)
 - C5.0(Quinlan, 1998)
 - 이들의 장점을 결합한 다양한 알고리즘

ID3 알고리즘

- 대표적인 의사결정트리 기반 분류 알고리즘
 1. 전체 데이터를 포함하는 루트 노드를 생성한다.
 2. 만약 샘플들이 모두 같은 클래스라면, 노드는 Leaf 노드가 되고, 해당 클래스로 레이블을 부여한다. (종료 조건)
 3. 그렇지 않으면 정보이득(Information Gain)이 높은 (즉, 데이터를 가장 잘 구분할 수 있는) 속성을 선택한다. (이때 정보이득은 엔트로피(Entropy)의 변화를 가지고 계산한다.)
 4. 선택된 속성으로 가지(Branch)를 뺀 하위 노드들을 생성한다. (각 하위 노드들은 가지의 조건을 만족하는 레코드들이다.)
 5. 각 노드에 대하여 2단계로 이동한다.

ID3 알고리즘 (예제를 통한 이해)

- 대표적인 의사결정트리 기반 분류 알고리즘
- 학습 예제를 가장 잘 분류할 수 있는 속성을 루트에 둬
· 엔트로피(Entropy), 정보 이득(Information Gain)
- 엔트로피 : 주어진 데이터 집합의 혼잡도를 의미
 $Entropy(S) = - P(+) * \log_2 (P(+)) - P(-) * \log_2 (P(-))$
 $P(+)$: 긍정 레코드의 개수 / 총 레코드의 개수
 $P(-)$: 부정 레코드의 개수 / 총 레코드의 개수

Entropy(S) 구하기

$$S = PlayTennis[9+, 5-]$$

$$P(+) = \frac{9}{14}$$

$$P(-) = \frac{5}{14}$$

$$Entropy(S) = - \frac{9}{14} \times \log_2 \frac{9}{14} - \frac{5}{14} \times \log_2 \frac{5}{14} = 0.940$$

ID3 알고리즘

- 정보 이득 : 데이터를 분할하기 전과 후의 변화
앞서 구한 Label Entropy인 Entropy(S)와
Node의 Entropy의 차이를 구함

$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

1. Gain(S, Outlook) 구하기

Values(Outlook) = Sunny, Overcast, Rain

$S_{Sunny} = [2+, 3-]$, $S_{Overcast} = [4+, 0-]$, $S_{Rain} = [3+, 2-]$

Outlook 노드에서 Rain이라는 값을 가진 레코드는 총 5개이고 5개 중에서 PlayTennis가 Yes인 경우가 3개, No인 경우가 2개

$$Gain(S, Outlook) = Entropy(S) - \sum_{v \in Sunny, Overcast, Rain} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$= Entropy(S) - \frac{5}{14} \times Entropy(S_{Sunny}) - \frac{4}{14} \times Entropy(S_{Overcast}) - \frac{5}{14} \times Entropy(S_{Rain})$$

$$= 0.940 - \frac{5}{14} \times 0.971 - \frac{4}{14} \times 0 - \frac{5}{14} \times 0.971$$

$$= 0.246$$

ID3 알고리즘

2. Gain(S, Wind) 구하기

$Values(Wind) = Weak, Strong$

$S_{Weak} = [6+, 2-]$

$S_{Strong} = [3+, 3-]$

$$\begin{aligned} Gain(S, Wind) &= Entropy(S) - \sum_{v \in \{Weak, Strong\}} \frac{S_v}{S} Entropy(S_v) \\ &= Entropy(S) - \frac{8}{14} \times Entropy(S_{Weak}) - \frac{6}{14} \times Entropy(S_{Strong}) \\ &= 0.940 - \frac{8}{14} \times 0.811 - \frac{6}{14} \times 1.00 \\ &= 0.048 \end{aligned}$$

ID3 알고리즘

3. Gain(S, Humidity) 구하기

$Values(Humidity) = High, Normal$

$S_{High} = [3+, 4-]$

$S_{Normal} = [6+, 1-]$

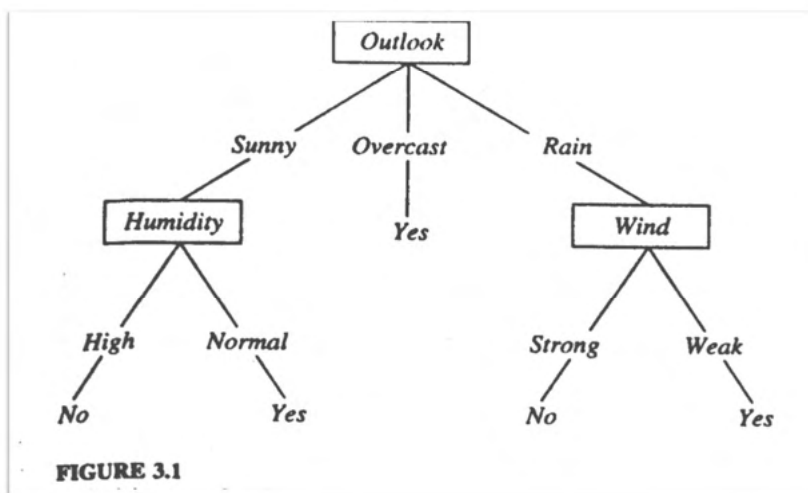
$$\begin{aligned} Gain(S, Humidity) &= Entropy(S) - \sum_{v \in \{High, Normal\}} \frac{S_v}{S} Entropy(S_v) \\ &= Entropy(S) - \frac{7}{14} \times Entropy(S_{High}) - \frac{7}{14} \times Entropy(S_{Normal}) \\ &= 0.940 - \frac{7}{14} \times 0.985 - \frac{7}{14} \times 0.592 \\ &= 0.151 \end{aligned}$$

ID3 알고리즘

Outlook 노드의 Information Gain : 0.246
Wind 노드의 Information Gain : 0.048
Humidity 노드의 Information Gain : 0.151

따라서 Information Gain이 가장 높은
"Outlook" 노드를 Root Node로 선택

완성된 트리 사례



예제 프로그램

화장품 구매여부 예측하기

	A	B	C	D	E	F	G
1	고객번호	성별	나이	직장여부	결혼여부	차량보유여부	구매여부
2	1	남	30대	NO	YES	NO	NO
3	2	여	20대	YES	YES	YES	NO
4	3	여	20대	YES	YES	NO	NO
5	4	여	40대	NO	NO	NO	NO
6	5	여	30대	NO	YES	NO	NO
7	6	여	30대	NO	NO	YES	NO
8	7	여	20대	NO	YES	NO	NO
9	8	여	20대	NO	YES	YES	YES
10	9	여	30대	YES	YES	NO	YES
11	10	남	40대	YES	NO	YES	NO
12	11	남	20대	NO	NO	NO	NO
13	12	남	30대	NO	YES	YES	NO
14	13	남	20대	YES	NO	NO	NO
15	14	여	30대	YES	YES	NO	YES
16	15	남	30대	YES	YES	YES	YES
17	16	여	30대	YES	NO	NO	NO
18	17	여	30대	NO	YES	YES	YES
19	18	남	20대	YES	YES	NO	NO
20	19	남	40대	YES	NO	YES	NO

예제 프로그램

고객번호	성별	나이	직장여부	결혼여부	차량보유여부
31	남	30대	NO	YES	NO

Test Data



고객번호	성별	나이	직장여부	결혼여부	차량보유여부	구매여부
1	남	30대	NO	YES	NO	NO
2	여	20대	YES	YES	YES	NO
3	여	20대	YES	YES	NO	NO
4	여	40대	NO	NO	NO	NO
5	여	30대	NO	YES	NO	NO
6	여	30대	NO	NO	YES	NO
7	여	20대	NO	YES	NO	NO
8	여	20대	NO	YES	YES	YES
9	여	30대	YES	YES	NO	YES
10	남	40대	YES	NO	YES	NO
11	남	20대	NO	NO	NO	NO
12	남	30대	NO	YES	YES	NO
13	남	20대	YES	NO	NO	NO
14	여	30대	YES	YES	NO	YES
15	남	30대	YES	YES	YES	YES
16	여	30대	YES	NO	NO	NO
17	여	30대	NO	YES	YES	YES
18	남	20대	YES	YES	NO	NO
19	남	40대	YES	NO	YES	NO
20	여	40대	YES	YES	NO	YES
21	여	20대	NO	YES	YES	YES
22	남	30대	NO	NO	NO	NO
23	여	30대	YES	YES	NO	YES
24	남	30대	YES	NO	YES	NO
25	여	40대	NO	YES	YES	YES
26	남	30대	NO	YES	NO	NO
27	여	30대	YES	YES	YES	YES
28	여	40대	YES	NO	YES	NO
29	남	40대	YES	YES	NO	YES
30	여	40대	YES	YES	NO	YES

정보 이득이 가장 높은 "결혼여부" 노드를 루트로 선택



전체 Entropy : 0.9709505944546606
 성별의 InformationGain : 0.1162959898239365
 나이의 InformationGain : 0.02816996444958253
 직장여부의 InformationGain : 0.019819440750390438
결혼여부의 InformationGain : 0.32365019815155627
 차량보유여부의 InformationGain : 0.008690515487248751
 선택된 노드 : 결혼여부
 값 : YES

예제 프로그램

고객번호	성별	나이	직장여부	결혼여부	차량보유여부	구매여부
1	남	30대	NO	YES	NO	NO
2	여	20대	YES	YES	YES	NO
3	여	20대	YES	YES	NO	NO
5	여	30대	NO	YES	NO	NO
7	여	20대	NO	YES	NO	NO
8	여	20대	NO	YES	YES	YES
9	여	30대	YES	YES	NO	YES
12	남	30대	NO	YES	YES	NO
14	여	30대	YES	YES	NO	YES
15	남	30대	YES	YES	YES	YES
17	여	30대	NO	YES	YES	YES
18	남	20대	YES	YES	NO	NO
20	여	40대	YES	YES	NO	YES
21	여	20대	NO	YES	YES	YES
23	여	30대	YES	YES	NO	YES
25	여	40대	NO	YES	YES	YES
26	남	30대	NO	YES	NO	NO
27	여	30대	YES	YES	YES	YES
29	남	40대	YES	YES	NO	YES
30	여	40대	YES	YES	NO	YES

원래 Entropy : 0.970950594546686
 성별의 InformationGain : 0.0912774462416801
 나이의 InformationGain : 0.2099865470109874
 직장여부의 InformationGain : 0.06002335238512835
 결혼여부의 InformationGain : 0.0
 차량보유여부의 InformationGain : 0.0464393446710154
 선택된 노드 : 나이
 값 : 30대



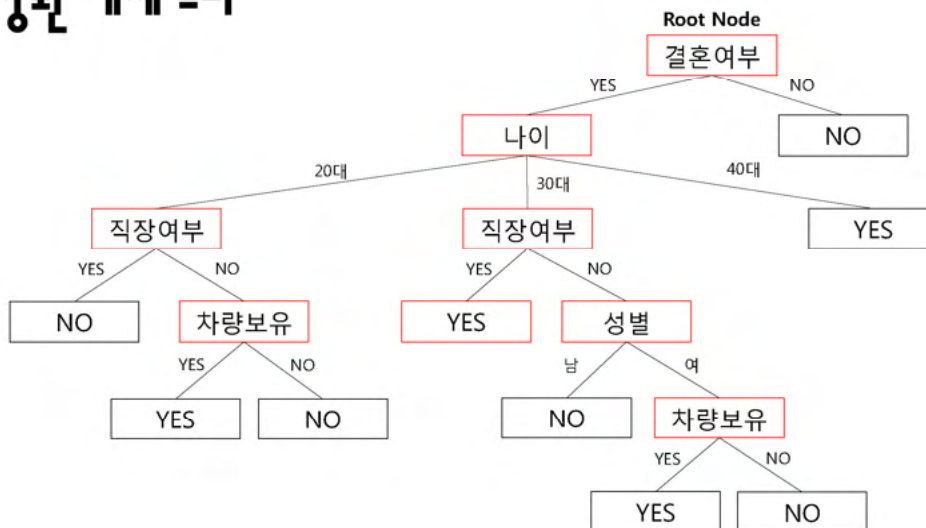
정보 이득이 가장 높은 "나이"를 다음 노드를 루트로 선택
 Entropy와 Information Gain을 계산해가며 "구매 여부"라는 Label이 같을 때 까지 반복


고객번호	성별	나이	직장여부	결혼여부	차량보유여부	구매여부
1	남	30대	NO	YES	NO	NO
12	남	30대	NO	YES	YES	NO
26	남	30대	NO	YES	NO	NO

분류 결과 : NO

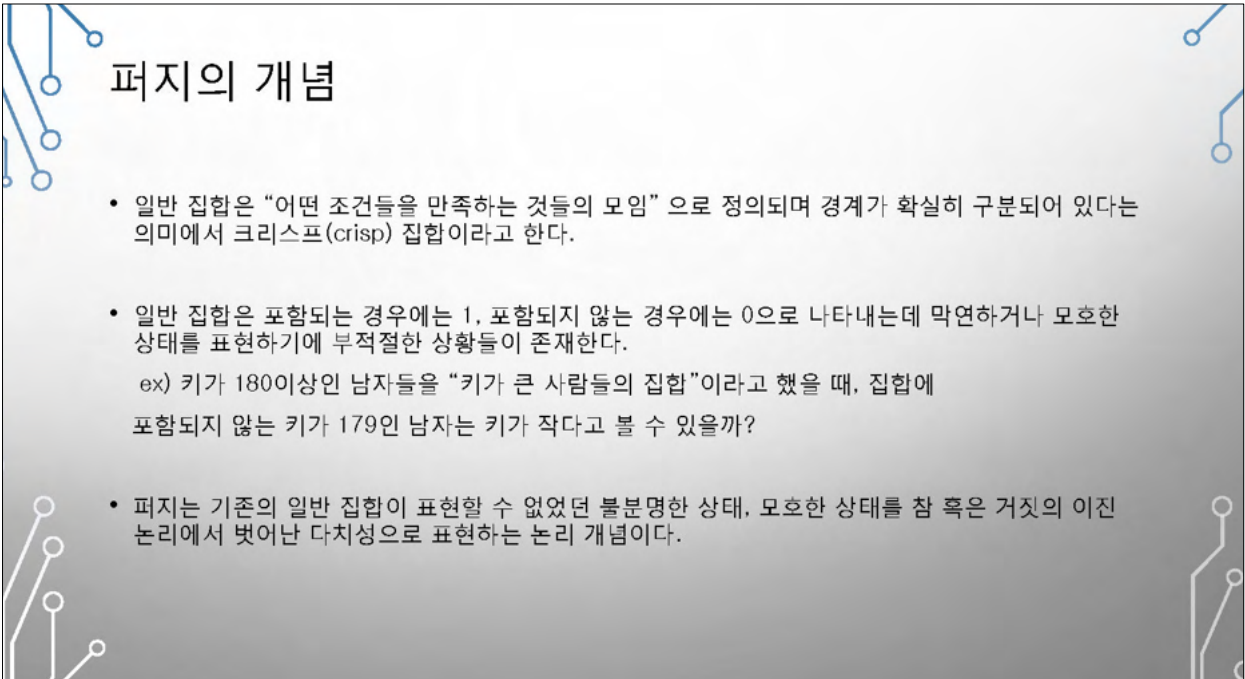
최종적으로 "NO"라는 분류 결과를 통해 Test Data는 구매하지 않는다고 예측

완성된 예제 트리





퍼지 의사결정트리



퍼지의 개념

- 일반 집합은 “어떤 조건들을 만족하는 것들의 모임”으로 정의되며 경계가 확실히 구분되어 있다는 의미에서 크리스프(crisp) 집합이라고 한다.
- 일반 집합은 포함되는 경우에는 1, 포함되지 않는 경우에는 0으로 나타내는데 막연하거나 모호한 상태를 표현하기에 부적절한 상황들이 존재한다.
ex) 키가 180이상인 남자들을 “키가 큰 사람들의 집합”이라고 했을 때, 집합에 포함되지 않는 키가 179인 남자는 키가 작다고 볼 수 있을까?
- 퍼지는 기존의 일반 집합이 표현할 수 없었던 불분명한 상태, 모호한 상태를 참 혹은 거짓의 이진 논리에서 벗어난 다치성으로 표현하는 논리 개념이다.

퍼지 집합

- 퍼지 집합의 명제는 참 또는 거짓이 아니라 부분적으로 참, 부분적으로 거짓이기 때문에 모호한 개념을 표현할 수 있다.
- 부분의 정도는 "소속도" 라고 하여 보통 [0,1] 범위의 실수값으로 표현한다.
- "키가 큰 남자" 라는 퍼지 집합의 원소는 모든 남자지만, 이 집합의 소속도는 표에서 보듯 키에 좌우된다.
- 키가 205cm인 마크의 소속도는 1, 키가 152cm인 피터의 소속도는 0이고 그 사이에 있는 사람들은 1과 0사이의 소속도를 가지며, 부분적으로 크고 작다고 볼 수 있다.

이름	키(CM)	소속도	
		크리스프	퍼지
크리스	208	1	1.00
마크	205	1	1.00
존	198	1	0.98
톰	181	1	0.82
데이비드	179	0	0.78
마이클	172	0	0.24
밥	167	0	0.15
스티븐	158	0	0.06
빌	155	0	0.01
피터	152	0	0.00

퍼지 집합

- 퍼지 집합은 어떤 요소가 그 집합에 어느 정도 포함되는지 그 소속도를 나타내기 위해 소속 함수(Membership Function)을 사용한다.
- 소속 함수는 집합의 안쪽을 1, 바깥쪽을 0으로 하고 경계에서는 그 사이 값을 취하여 안쪽에 가까울수록 1에 가까운 값을 가진다.
- 퍼지 집합은 소속 함수 μ 를 사용하여 다음과 같이 정의된다.

$\mu_A(x) \rightarrow [0,1]$

$\mu_A(x) = 1$	x가 완전히 A에 속한 경우
$\mu_A(x) = 0$	x가 완전히 A에 속하지 않은 경우
$0 < \mu_A(x) < 1$	x가 부분적으로 A에 속한 경우

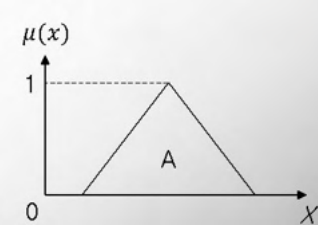
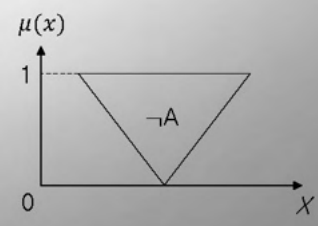
퍼지 집합 연산

- 여집합(Complement)
 - 크리스프 집합 : 어떤 원소가 그 집합에 속하지 않을까?
 - 퍼지 집합 : 어떤 원소가 그 집합에 "얼마만큼" 속하지 않을까?

$$\mu_{\neg A}(x) = 1 - \mu_A(x)$$

키가 큰 남자의 퍼지 집합 = (0/180, 0.25/182.5, 0.5/185, 0.75/187.5, 1/190)

키가 크지 않은 남자의 퍼지 집합 = (1/180, 0.75/182.5, 0.5/185, 0.25/187.5, 0/190)

퍼지 집합 연산

- 교집합(Intersection)
 - 크리스프 집합 : 어느 원소가 두 집합에 모두 속할까?
 - 퍼지 집합 : 원소가 두 집합 모두에 얼마만큼 속할까?

크리스프 집합의 교집합은 두 집합이 겹치는 영역을 말하지만, 퍼지 집합은 한 원소가 두 집합에 대해 서로 다른 정도로 속하기 때문에 각 원소의 퍼지 교집합에 대한 소속값은 두 집합에 대한 소속값 중 낮은 값이 된다.

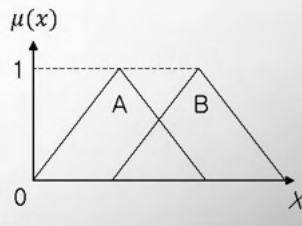
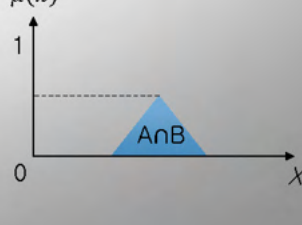
$$\mu_{A \cap B}(x) = \min[\mu_A(x), \mu_B(x)] = \mu_A(x) \cap \mu_B(x)$$

키가 큰 남자의 퍼지 집합 = (0/165, 0/175, 0/180, 0.25/182.5, 0.5/185, 1/190)

키가 보통인 남자의 퍼지 집합 = (0/165, 1/175, 0.5/180, 0.25/182.5, 0/185, 0/190)

키가 큰 남자 \cap 키가 보통인 남자 = (0/165, 0/175, 0/180, 0.25/182.5, 0/185, 0/190)

= (0/180, 0.25/182.5, 0/185)

퍼지 집합 연산

- 합집합(Union)

크리스프 집합 : 원소가 두 집합에 어느 쪽이든 속할까?

퍼지 집합 : 원소가 두 집합 어느 쪽이든 얼마만큼 속할까?

크리스프 집합의 합집합은 두 집합의 전체 원소를 포함하지만, 퍼지 집합의 합집합은 교집합의 반대이다.

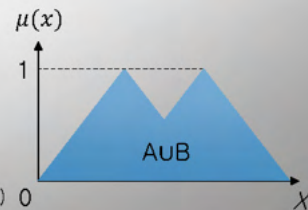
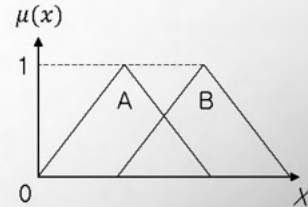
즉, 각 원소의 합집합에 대한 소속값은 두 집합에 대한 소속값 중 높은 값이다.

$$\mu_{A \cup B}(x) = \max[\mu_A(x), \mu_B(x)] = \mu_A(x) \cup \mu_B(x)$$

키가 큰 남자의 퍼지 집합 = (0/165, 0/175, 0/180, 0.25/182.5, 0.5/185, 1/190)

키가 보통인 남자의 퍼지 집합 = (0/165, 1/175, 0.5/180, 0.25/182.5, 0/185, 0/190)

키가 큰 남자 \cup 키가 보통인 남자 = (0/165, 1/175, 0.5/180, 0.25/182.5, 0.5/185, 1/190)



퍼지 의사결정트리

- 퍼지 의사결정트리는 기존의 의사결정트리에서 확장된 방법이며 불확실한 분류 문제에서 지식을 추출하는 효과적인 방법
- 퍼지 이론을 적용하여 데이터 집합을 표현하고 트리 구조를 결정
- 기존의 의사결정트리와 동일한 접근법을 사용
 - : 리프 노드에 도달하거나 속성 또는 레코드가 남아 있지 않을 때 까지 반복
- 분할 과정에서 퍼지 엔트로피와 퍼지 데이터 세트의 정보 이득을 계산하여 트리를 확장하기 위해 트리의 테스트 노드에서 사용할 속성을 선택

퍼지 의사결정트리

- 기존의 의사결정트리의 특징은 각 레코드가 특정 노드에 대해서 확실하게 속해 있거나 속해 있지 않지만 퍼지의 경우에는 다름
- 각 속성에 대해 언어변수를 정의하고, 주어진 예제의 소속도를 결정해야 함

온도
$x_1 = 8$
$x_2 = 10$
$x_3 = 15$
$x_4 = 20$
$x_5 = 23$
$x_6 = 25$

➔

온도		
시원함	화창함	더움
0.7	0.3	0.0
0.5	0.5	0.0
0.0	1.0	0.0
0.0	0.8	0.2
0.0	0.5	0.5
0.0	0.0	1.0

예제

- 퍼지 의사결정트리 구축을 위한 트레이닝 셋

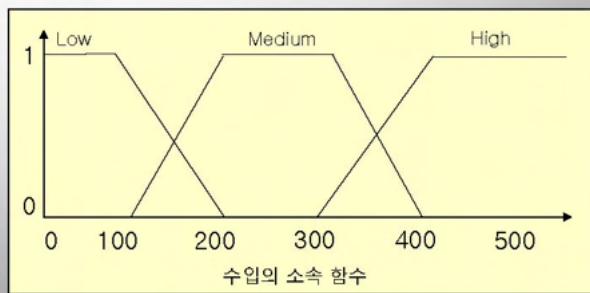
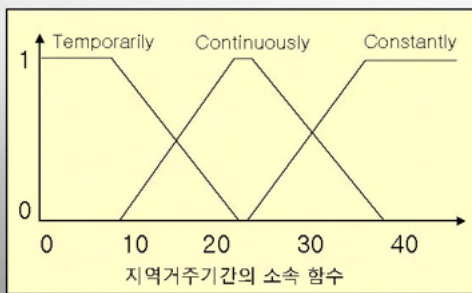
No.	지역거주기간(년)	수입(만원)	신용도
D1	1	100	0.0
D2	10	150	0.0
D3	15	200	0.1
D4	20	300	0.3
D5	30	250	0.7
D6	40	350	0.9
D7	40	500	1.0

- 지역 거주 기간이 25년이며 수입이 320만원인 고객이 은행으로 부터 대출을 받을 수 있을지에 대한 문제를 퍼지 의사결정 트리를 통해 분류

예제

• 속성에 대한 언어변수:

- “지역거주기간”은 “temporarily”, “continuously”, “constantly”;
- “수입”은 “Low”, “medium”, “high”.



트레이닝 셋의 퍼지화

• 소속 함수를 통해 변경된 속성들의 소속도

No.	지역거주기간			수입		
	Temporarily	Continuously	Constantly	Low	Medium	High
D1	1	0	0	1	0	0
D2	0.8	0.2	0	0.6	0.4	0
D3	0.5	0.5	0	0.1	0.9	0
D4	0.2	0.8	0	0	1	0
D5	0	0.5	0.5	0	1	0
D6	0	0	1	0	0.6	0.4
D7	0	0	1	0	0	1

퍼지 의사결정트리 구현

1. 총 엔트로피 구하기

- $P_{yes} = 0 + 0 + 0.1 + 0.3 + 0.7 + 0.9 + 1.0 = 3$ (“신용도” 속성의 *positive examples*)
- $P_{no} = 1 + 1 + 0.9 + 0.7 + 0.3 + 0.1 + 0 = 4$ (“신용도” 속성의 *negative examples, P_{yes} 의 여집합*)
- $P = P_{yes} + P_{no} = 3 + 4 = 7$ (*all examples*)

아래 식을 통해 전체 엔트로피 $E(S^N)$ 을 계산 :

$$E(S^N) = -\frac{3}{7} \cdot \log_2 \frac{3}{7} - \frac{4}{7} \cdot \log_2 \frac{4}{7} \approx 0,985 \text{ bit.}$$

퍼지 의사결정트리 구현

2. 지역거주기간 속성의 temporarily에 대한 엔트로피 구하기

$E(S^N, \text{지역거주기간})$:

- $P_{yes}^{\text{temporarily}} = \min(0, 1) + \min(0; 0.8) + \min(0.1; 0.5) + \min(0.3; 0.2) + \min(0.7, 0) + \min(0.9; 0) + \min(1, 0) = 0 + 0 + 0.1 + 0.2 + 0 + 0 + 0 = 0.3$
- $P_{no}^{\text{temporarily}} = \min(1, 1) + \min(1; 0.8) + \min(0.9; 0.5) + \min(0.7; 0.2) + \min(0.3, 0) + \min(0.1; 0) + \min(0, 0) = 1 + 0.8 + 0.5 + 0.2 + 0 + 0 + 0 = 2.5$
- $P^{\text{temporarily}} = 0.3 + 2.5 = 2.8$

따라서, $E(\text{지역거주기간, temporarily}) = -\frac{0,3}{2,8} \cdot \log_2 \frac{0,3}{2,8} - \frac{2,5}{2,8} \cdot \log_2 \frac{2,5}{2,8} \approx 0,491 \text{ bit.}$

퍼지 의사결정트리 구현

- Temporarily와 마찬가지로 Continuously, Constantly도 동일한 계산을 수행하면 아래 표와 같은 결과를 얻을 수 있음

	Temporarily	Continuously	Constantly
P_{yes}	0.3	0.9	2.4
P_{no}	2.5	1.7	0.4
E (bit)	0.491	0.931	0.592

퍼지 의사결정트리 구현

3. “지역거주기간” 속성의 총 엔트로피 계산:

- $E(S^N, \text{지역거주기간}) = \frac{2.5}{7} \cdot 0.491 + \frac{1.7}{7} \cdot 0.931 + \frac{0.4}{7} \cdot 0.592 \approx 0.486 \text{ bit}$

4. “지역거주기간” 속성에 대한 “정보 이득” 계산: <총 엔트로피 - 속성 엔트로피>

- $G(S^N, \text{지역거주기간}) = 0.985 - 0.486 = 0.499 \text{ bit}$

“수입” 속성에 대해서도 동일한 계산을 수행:

- $E(S^N, \text{수입}) = 0.416 \text{ bit}$

- $G(S^N, \text{수입}) = 0.569 \text{ bit}$

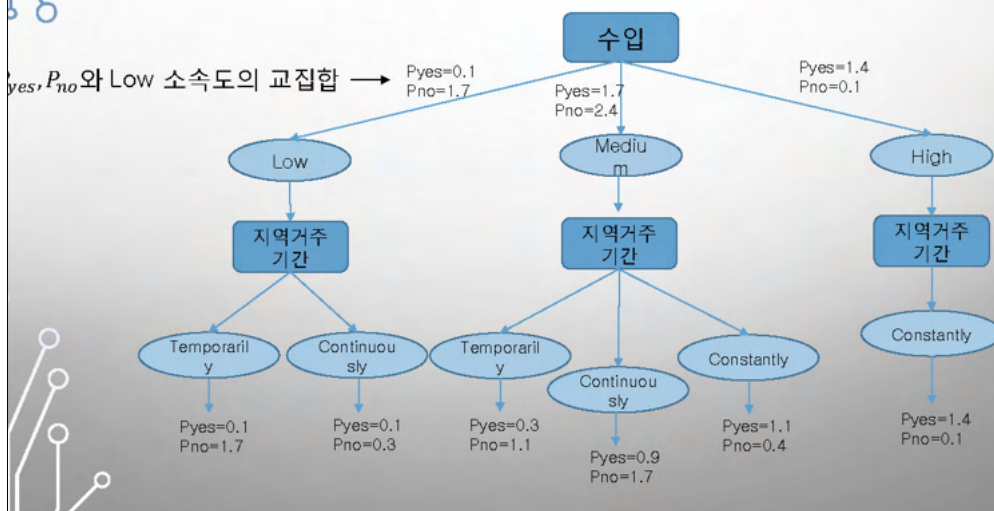
정보 이득 계산 결과 “수입” 속성의 정보 이득이 가장 높으므로 “수입” 속성을 트리의 루트 노드로 선택

5. 수입의 속성에 대한 각 지역거주기간 속성들의 소속도를 계산

- 수입(Low,Medium,High)과 지역거주기간(Temp,Cont,Const)의 소속도의 교집합
- 소속도가 모두 0인 속성은 제거

수입	Low			Medium			High		
	Temporarily	Continuously	Constantly	Temporarily	Continuously	Constantly	Temporarily	Continuously	Constantly
D1	1	0	0	0	0	0	0	0	0
D2	0.6	0.2	0	0.4	0.2	0	0	0	0
D3	0.1	0.1	0	0.5	0.5	0	0	0	0
D4	0	0	0	0.2	0.8	0	0	0	0
D5	0	0	0	0	0.5	0.5	0	0	0
D6	0	0	0	0	0	0.6	0	0	0.4
D7	0	0	0	0	0	0	0	0	1

구현된 퍼지 의사결정트리의 모습



- 노드 [지역거주기간 = temporarily and 수입 = high]
 [지역거주기간 = continuously and 수입 = high]
 [지역거주기간 = constantly and 수입 = low]
 들은 레코드의 소속도가 모두 0이므로 트리에서 제거
- 완성된 트리를 기반으로 지역거주기간이 25년이고 수입이 320만원인 신규 고객의 신용 등급을 정의

- 신규 고객은 두 영역에 속함: [지역거주기간 = continuously and 수입 = medium]
 [지역거주기간 = constantly and 수입 = high]
- 소속 함수에 따른 신규 고객의 소속도는 각각 [0.8, 0.8] / [0.2, 0.2] = $\mu_1(D_j)$
- 트리를 따라 얻은 P_{yes} 와 P_{no} 의 값은 각각 [0.9, 1.7] 과 [1.1, 0.4] = P_k^i
- 신규 고객에 대한 신용도를 구하기 위해 아래와 같은 식을 사용

$$\sigma_j = \frac{\sum_i \sum_k P_k^i \cdot \mu_1(D_j) \cdot x_k}{\sum_i (\mu_1(D_j) \cdot \sum_k P_k^i)}, \quad (7)$$

$x_{yes} = 1, x_{no} = 0$

$$\therefore \frac{0.9 * 0.8 * 1 + 1.7 * 0.8 * 0 + 1.4 * 0.2 * 1 + 0.1 * 0.2 * 0}{(0.9 + 1.7) * 0.8 + (1.4 + 0.1) * 0.2} = 0.420$$

- 신규 고객은 0.420 이라는 신용 등급을 받았으며, 대출을 받지 못함

결언

- 의사결정트리 모형은 데이터마이닝의 주요기법으로 자리잡고 있으며 SAS/EMINER, SPSS AnswerTree, R, Weka 등에서 이를 사용
- 의사결정트리는 예측하거나 분류하는 쉽고 투명한 방법
- 의사결정트리는 데이터 구조에 대한 어떠한 가정을 갖지 않아, 다량의 샘플 필요
- 의사결정트리는 결과를 해석하고 이해하기에 쉽고, 자료를 가공할 필요가 거의 없는 방법
- 최적의 결정 트리를 학습하는 문제는 NP-complete 문제로 알려져 있으며, 최적 결정 트리를 알아낸다고 보장 못함
- 퍼지의사결정트리는 각 특징값들이 경계면에 가까이 있는 경우 발생할 수 있는 분류 오류의 가능성을 줄일 수 있는 방법

감사합니다 ...

※ 저작권법에 의해 무단전재와 무단복제를 금합니다.
※ 본 권의 모든 저작권은 한국정보통신학회에 있습니다.

·이 책은 부산인재평생교육진흥원의 일부 재정지원에 의하여 발간되었음.

한국정보통신학회 4차 산업 관련 인공지능 기반 기술 워크숍

서 기 2018년 8월 21일 인쇄

서 기 2018년 8월 24일 발행

발행처 : 사단
법인 한국정보통신학회

부산시 부산진구 서면문화로 27, 1802호(부전동, 유원오피스텔)

전 화 : 051)463-3683

팩 스 : 051)464-3683

홈페이지 : <http://www.kiice.org> Email : kiice@kiice.org

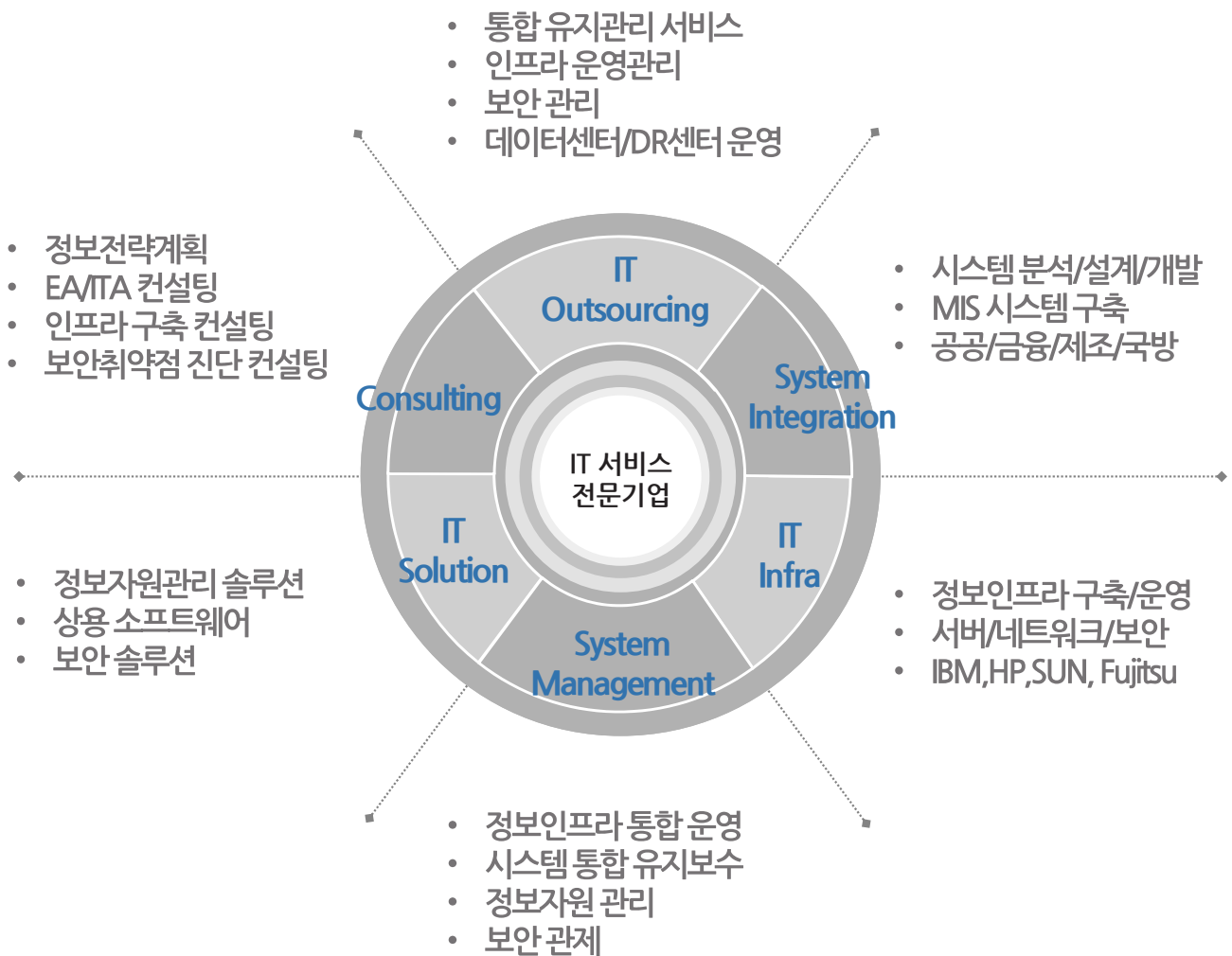
조판·인쇄소 : **한림원(주)** TEL : (02) 2273-4201 <http://www.hanrimwon.com>

사단
법인 **한국정보통신학회**

4차 산업 관련 인공지능 기반 기술 워크숍 개최를 축하합니다.

주식회사 록스

정보시스템 구축 및 통합 유지관리 전문기업 (주)엠티데이터 는
공공기관 및 대기업 유지관리 사업의 축적된 Know-How와 전문 기술인력을 바탕으로
최적의 시스템 구축 및 운영.유지관리 서비스를 제공합니다.



고객과 생사고락을 함께하는 최고의 IT파트너

(주)엠티데이터

본 사 : 경기도 성남시 분당구 판교로255번길9-22,514호(삼평동,우림W-CITY)
대전지사 : 대전시 유성구 죽동로 69 소프트아이텍 2층(죽동)
부산지사 : 부산시 남구 수영로 309 경성대 사회관 B106호(대연동)
Tel : 031-628-8180 Fax : 031-628-8188 www.mtdata.co.kr

아이씨티웨이

정보통신기술벤처기업

We design optimal way for ICT

IT와 관련된 특별한 서비스를 원하십니까?

저희는 **ICT Infrastructure**, **Specialized Consulting**, **System Integration**, **Total Outsourcing**에 대한 최적의 서비스를 제공합니다.

20년 이상의 다양한 프로젝트 경험과 분야별 전문기술 인력을 보유하고 있는 ICTWAY와 상의하십시오.

만족을 넘어 감동을 안겨 드리겠습니다.



Information

Communication

Technology

사물인터넷(IoT) 교육을 위한 삼성 ARTIK 실습용 교육 키트 + 교재 (ARTIK 710, 053, 530)



교육키트
구성

ARTIK
교재



1층 서랍



2층 서랍



3층 서랍



4층 서랍



<ARTIK 053, 530, 710 교재 예제 실습 가능>



과학상자 시리즈



MDF 시리즈



- ARTIK 교육 키트는 총 4개의 서랍 및 펜트하우스로 구성
- 각 서랍은 교재의 예제 실습을 위해 IoT 확장 모듈 37종과 USB 카메라, LED 모듈 등의 액츄레이터로 구성
- 펜트하우스는 "메이커스 시리즈(과학상자, MDF)" 및 모터 드라이버, 브레드 보드로 구성

- ARTIK 및 IoT의 기본 개념 정리
- 교육 키트의 센서, IoT 확장 모듈의 원리, 사용법, 소스 코드 등의 내용 수록
- 기본 예제 및 응용 예제를 통한 센서 학습ARTIK
- 초보자부터 S/W 개발자까지 실습 가능한 레벨로 구성
- 30여 개의 다양한 실습 및 기본/응용 예제 수록
- 교육이론, 실습 커리큘럼 제공



DIY

전용앱
제공

ARTIK IoT 공기청정기 Kit

- ARTIK IoT 메이커스 시리즈 중 하나로, 공기청정기와 IoT 디바이스를 조립하여 직접 만들어 볼 수 있는 교육용 콘텐츠
- 아크릴 외부 프레임 만들고 ARTIK 053을 이용하여 센서 제어를 통해 공기청정기 제작
- ARTIK 053을 AP로 동작 시킨 후 TCP 소켓 통신을 이용하여 앱을 통한 모니터링 구현
- 앱의 기능 : 미세먼지 수치, 공기청정기 On/Off, 미세먼지 수치 모니터링을 통한 자동 작동 기능

환경과 에너지를 생각하는 IoT 전문기업

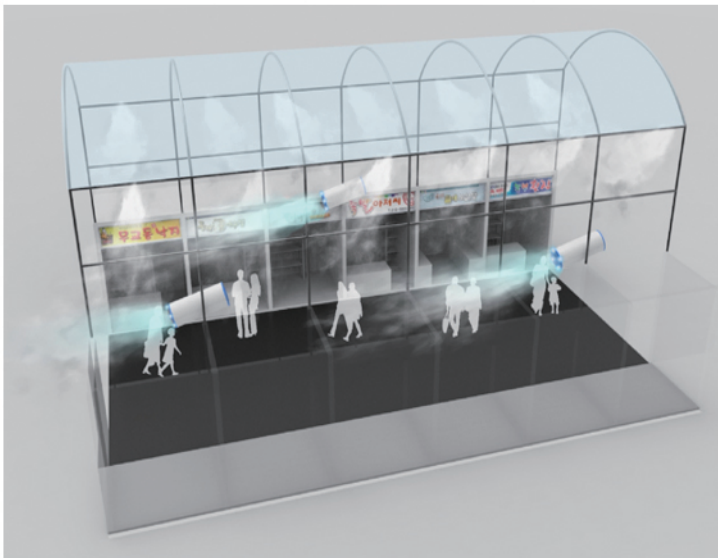
(주)혜안은 환경개선을 위한 미세 안개분무 시스템과 모터기반 시설 안전관리를 위한 IoT 스마트 모터 보호 시스템을 주력으로 하는 부산지역 IoT 전문기업입니다. TEL 051) 914- 3399

미세 안개분무 시스템

미세 안개분무 시스템은 구조의 변경 없이 설치가 가능하며, 효과적이고 친환경적으로 온도를 하락시키고 공기의 질을 개선할 수 있어 고객과 상인 모두에게 긍정적인 영향을 기대할 수 있습니다.

(주)혜안은 미세먼지 농도 증가에 따른 대기 오염과 여름철 더위에 취약하고 정체된 내부 공기 질을 개선시킬 수 있는 친환경적이고 효율적인 방법을 제공하고 있습니다.

서비스 개념도



시스템 구성



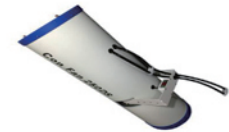
AWDS / AADS



Misting Pump



Multi Tube / Nozzle



CONFAN

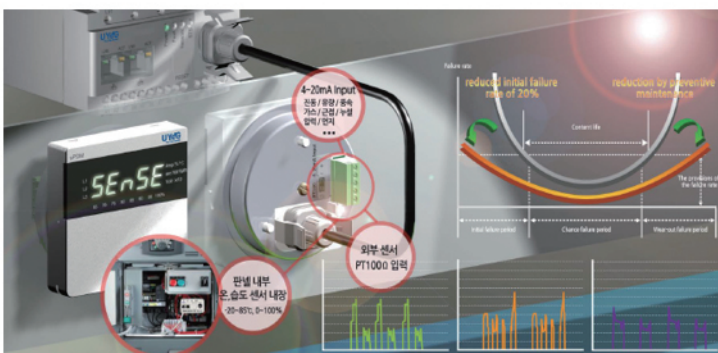
- 적용분야
- 전통시장, 농축산물 관리시설 등

IoT 스마트 모터 보호 시스템

예지보전이란 기기의 이상을 그 상태감시에 의하여 예지하고, 그 정보에 기인해서 행하는 보전하는 것을 의미합니다.

IoT 스마트 모터보호 계전기는 모터의 전류, 전압, 전력량, 팬벨 온/습도, 모터 온도를 한번에 보호 및 감시 할 수 있는 토털솔루션을 제공합니다.

서비스 개념도



시스템 구성



UYeG-ZE (IoT-EOCR)

- 적용분야
- 아파트, 빌딩, 산업시설, 공공시설 등

무한한 가능성이 만들어 내는

새로운 통신세상

고객의 행복을 추구하는 통신 세상의 새로운 리더!
고객님의 행복한 삶을 위하여 SK브로드밴드가 열어가겠습니다.



**SK broadband가 한국정보통신학회의
발전을 응원합니다.**

Since 1991 & future

www.hanscom.com

hans **com**

SINCE 1991 주식회사한스콤정보통신

정보통신 선도기업

네트워크통합(NI)

시스템통합(SI)

보안(Security)

IT 유지보수

통합커뮤니케이션(UC)

hans **com**

SINCE 1991 주식회사한스콤정보통신

부산광역시 동래구 여고북로56 (사직3동 358-6 한빌딩 3층)

tel 051.507.0364 fax 051.507.0366

Global IT Leader!

모든 비즈니스 영역을 통합하는 통찰력으로
고객의 니즈를 완벽히 분석한 최적의 서비스로
미래를 선도하는 최첨단 기술력으로

미래의 가치를 먼저 생각하는 기업



Total Solutions

- SI-NI 사업
- Print On Demand 솔루션 사업



Smart Service

- Mobile 솔루션 사업
- 금융 솔루션 사업



Art Technologies

- 산업용 PDA 사업



큰 大 믿을 信

대신정보통신주식회사 Daishin Information & Communications Co., Ltd.

본사 : 광주광역시 서구 상무중앙로 110

서울 : 서울특별시 금천구 가산디지털1로, 205-28 대신정보통신빌딩

Tel_062-225-7350

Fax_062-226-0716

Tel_02-2017-5000

Fax_02-2107-5015

www.dsic.co.kr

GST

Smart Factory / ERP / MES / POP / IoT

SMART FACTORY

기업정보화의 핵심솔루션인

ERP, MES/POP, SCM, PMS, QMS, GROUPWARE 및

임베디드시스템을 중심으로 하는 Smart Factory 구축 및 E-Business 솔루션 전문기업

ERP시스템인 PLUS WIN 4.0은 중소중견기업의 경쟁력신형 정보화 시스템으로

동남·경제권역권의 조선기자재, 해양플랜트, 자동차부품, 섬유, 화학, 반도체, 무역업계 등 다양한 업종에 성공적으로 적용

ITCEN FLIES BEYOND INNOVATION

대한민국 정보기술의 중심, 혁신을 날다

