

날씨 마케팅 적용을 위한 기후 데이터의 군집 분석†

Cluster Analysis of Climate Data for Applying Weather Marketing

이양구*, 김원태**, 정영진***, 김광득****, 류근호*****

Yang-Koo Lee, Won-Tae Kim, Young-Jin Jung, Kwang-Deuk Kim, Keun-Ho Ryu

요약 최근 환경오염으로 인한 날씨의 변화, 자원 고갈에 따른 국제 유가의 상승 등 날씨 및 에너지 문제가 기업이나 국가 심지어 개인의 일상생활, 경제활동에 크나큰 영향을 미치고 있다. 이와 같은 이유로 대체에너지 중 태양 에너지 개발에 필요한 일사량 관리와 기후 데이터의 변화 특성 등을 근거로 지역성 규명에 관한 연구가 많이 이루어지고 있다. 그러나 아직까지는 데이터 마이닝을 이용한 지역적 특성에 따른 군집 및 체계적인 분석 데이터 검색 서비스가 효과적으로 제공되지 않고 있다. 따라서 이 논문에서는 국내에서 측정된 기후데이터를 저장 및 관리하기 위한 데이터를 모델링하고, k-means 기법을 이용하여 국내 기후 데이터를 지역적 특성에 따라 군집함으로써 체계적인 데이터 정보를 제공한다. 그리고 이러한 정보들이 날씨 마케팅에 어떻게 적용되는가에 대한 사례를 보인다. 제안 시스템은 기업의 날씨 마케팅 연구 및 이에 영향을 미치는 요소와 분석 정보를 제공할 수 있는 기본 데이터베이스 구축에 유용하게 활용될 것이다.

Abstract Recently, the weather has been influenced by the environmental pollution and the oil price has been risen because of the lack of resources. So, the weather and energy are influencing on not only enterprises or nations, but also individual daily life and economic activities very much. Because of these reasons, there are so many researches about management of solar radiation needed to develop solar energy as alternative energy. And many researchers are also interested in identifying the area according to changing characteristics of climate data. However, the researches have not developed how to apply the cluster analysis, retrieval and analytical results according to the characteristics of the area through data mining. In this paper, we design a data model of the data for storing and managing the climate data tested in twenty cities in the domestic area. And we provide the information according to the characteristics of the area after clustering the domestic climate data, using k-means clustering algorithm. And we suggest the way how to apply the department store and amusement park as an applied weather marketing. The proposed system is useful for constructing the database about the weather marketing and for providing the elements and analysis information.

주요어 : 군집 분석, 이력 데이터베이스, 기후 이력 데이터 관리, 날씨 마케팅

KeyWords : Cluster Analysis, History Database, Climate History Data Management, Weather Marketing

1. 서론

최근 환경오염으로 인한 날씨의 변화, 자원 고갈에 따른 국제 유가의 상승 등 날씨 및 에너지 문제가 기업이나 국가 심지어 개인의 일상생활, 경제활동에 크

나큰 영향을 미치고 있다. 이와 같은 이유로 대체에너지 중 태양 에너지 개발에 필요한 일사량 관리와 기후의 데이터 변화 특성 등을 근거로 하는 지역성 규명에 관한 연구가 많이 이루어지고 있다. 이러한 지역에 따른 기후 데이터 정보는 산업분야와 관련업종의 사전

† 이 연구는 산업자원부 한국산업기술 평가원 지정 청주대학교·정보통신연구센터 및 한국전자통신연구원(ETRI) 지원에 의한 것입니다.

* 충북대학교 대학원 전자계산학과 박사과정

** 충북대학교 대학원 전자계산학과

*** 충북대학교 대학원 전자계산학과 박사과정

**** 한국에너지기술연구원 책임기술원

***** 충북대학교 전기전자 및 컴퓨터공학부 교수

leeyangkoo@dblab.chungbuk.ac.kr

wtkim@dblab.chungbuk.ac.kr

yjjeong@dblab.chungbuk.ac.kr

kdkim@kier.re.kr

khryu@dblab.chungbuk.ac.kr

수요와 피해의 예측을 통한 이익극대화 및 손실최소화를 추구하는데 있어 매우 중요하다[1],[2],[3].

지금까지 국외에서는 기후 데이터에 대한 개발 및 보급이 활발히 추진되고 있으며, 이에 대한 정보를 인터넷으로 서비스하기 위해 많은 노력을 하고 있다. 또한 이런 정보를 분석하고 예측하기 위해 이들 데이터의 특성과 밀접한 시공간 데이터 마이닝에 대한 연구가 활발하게 진행되고 있다[1],[3]. 그러나 아직까지는 데이터 마이닝을 통한 지역적 특성에 따른 군집 및 체계적인 분석 데이터 검색 서비스가 효과적으로 제공되지 않고 있다.

이 논문에서는 국내 20개 도시에서 측정된 기후 데이터를 저장 및 관리하기 위한 데이터를 모델링하고, k-means 기법을 이용하여 국내 기후 데이터를 지역적 특성에 따라 군집함으로써 체계적인 데이터 정보를 제공한다. 그리고 이러한 정보들이 어떻게 날씨 마케팅에 적용될 수 있는가에 대한 응용 사례를 보인다. 제안 시스템은 기업의 날씨 마케팅 연구 및 이에 영향을 미치는 요소와 분석 정보를 제공할 수 있는 기본 데이터베이스 구축에 유용하게 활용될 것이다.

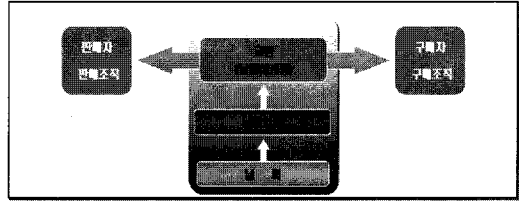
이 논문의 구성은 다음과 같다. 먼저, 2장에서는 관련 연구로 시공간 데이터 마이닝과 군집에 대해서 살펴본다. 3장에서는 관계형 데이터 모델을 기반으로 시간 속성을 추가함으로써 기후 데이터의 체계적인 저장 및 관리를 위한 기후 데이터 모델을 제시하고, 4장에서는 기후 데이터 분석을 위해 k-means[8] 기법의 적용 및 국내 기후 데이터 중 강수량을 이용한 지역 분석 결과를 제시한다. 5장에서는 제안 시스템의 설계 및 구현을 제시하고, 6장에서는 제안된 시스템이 어떻게 적용될 수 있는가에 대한 응용 사례를 보인다. 마지막으로 7장에서는 결론 및 향후 연구 방향을 제시한다.

2. 관련연구

2.1 날씨 마케팅의 정의 및 적용 분야

날씨 마케팅(weather marketing)이란 교환을 통한 인간의 욕구충족과 관련된 모든 활동을 마케팅이라 할 때, 날씨 마케팅 또는 기상 마케팅은 날씨를 마케팅에 직접 연계시키는 것이다[5],[6],[7]. 즉 날씨와 마케팅을 연관하여 교환 활성화와 시장 활성화를 도모하는 것이라고 정의할 수 있다. <그림 1>의 날씨 마케팅

의 개념적 도식에서 보는 바와 같이 날씨 마케팅은 판매자(조직), 교환과정, 그리고 구매자(조직)에 대한 날씨의 영향을 고려하거나 활용하는 것을 의미한다.



<그림 1> 날씨 마케팅의 개념적 도식

현대에 이르러 날씨가 기업이나 국가 심지어 개인 의 일상생활, 경제활동에 크나큰 영향을 미친다는 사실은 정설로 받아들여지고 있다. 날씨가 경제활동에 미치는 부정적인 영향은 전 세계적으로 30억 달러 이상이 추정되고 있으며 미국의 경우 70% 이상의 기업들이 날씨에 의해 매출이나 비용이 영향을 받고 있는 것으로 추정되고 있다. 예를 들어 <그림 2>에서와 같이 산업 기후 연구소에서 발표한 날씨와 유통업 매출과의 연관 관계는 매우 크다는 명확한 증거가 제시되고 있다. <그림 2>에서의 임계온도는 상품의 수요가 급격히 변하는 시점의 기온으로 판매율과 기온의 상관관계를 분석하는 기준이 된다.

관측종류	최고기온 20도가 넘으면 맑거나 시차. 25도가 넘으면 1도 상승할 때마다 판매량이 20% 증가
우유, 요구르트	기온 상승과 매출이 반비례하는 제품. 20도에서 30도로 기온이 상승할 때 매출은 약 8% 감소
쌀	25도를 넘으면 가격이 급격히 증가. 기온이 1도 올라갈 때마다 판매량이 15% 증가
켄 케피	기온이 30도 가까이 되면 갑자기 잘 팔린다. 25도 넘어서면 1도 상승할 때마다 판매량이 18% 증가
스포츠 드링크	최고기온 25도부터 매출이 높아짐. 30도 가까이 되면 판매량이 급격히 높아진다. 25도를 넘으면 2도 상승할 때마다 8%씩 매출 증가
아이스크림과 방과후	25~30도에서는 유 지분이 든 아이스크림. 30도를 넘으면 얼음이 많은 방과후가 잘 팔린다. 20도 이하로 낮아지면 소파가 감소
찬시국수	최고기온이 25도가 될 때부터 잘 팔리기 시작. 30도를 전후하여 절정을 이룬다. 25도가 넘으면 1도 상승할 때 이다 15%의 비율로 매출 증가
콜드와 데니스	평균기온 18도 이하부터 저는 시량이 늘어나 24도에서 절정을 이룬다. 27도를 넘으면 콜드 데니스를 팔기는 시량이 감소
신시복	최고기온이 28도가 되면 20% 증가. 낮 기온이 27도 이하로 떨어지고, 아침기온이 20도 이하로 내려 기온 상승에 따른 매출이 80% 이상 증가

<그림 2> 날씨와 유통업 매출간의 연관 관계 -임계온도

2.2 해외 날씨 마케팅의 사례

해외의 경우 기상을 상업화하는 일본, 미국 등 선진국에서는 일찍이 고부 가치의 날씨 마케팅을 개발하여 유료로 서비스하는 것이 일반화되어 있다. 미국의 경우에는 기상 마케팅의 원조국이라는 이미지에 걸맞게 다양한 민간 서비스가 실시되고 있다. 가장 유명한 민

간 서비스 업자 중의 하나는 날씨 뉴스(weather news)사이다. 날씨 뉴스사의 경우 인터넷을 활용하여 세계 각지의 구체적인 기상 정보를 제공하고 있다. 예를 들어 만약 내가 유럽을 여행하고 싶다면 이 회사의 홈페이지에서 내가 여행할 구체적인 각국의 도시들에 관한 기상 정보를 나의 여행 일정별로 미리 알 수 있다.

미국의 유명한 대형 유통점인 K마트의 경우는 기상 정보를 잘 활용하고 있는 대표적인 예이다. K마트의 경우에는 잔디 씨앗부터 제설기까지 다양한 제품을 팔고 있는데 각 점포에 물건을 배송해주는 책임자의 임무는 필요한 물건을 제때 적소에 배달해주는 것이다. 여기서 중요한 것은 수요를 미리 예측하고 이에 대응해야 한다는 것이다. 이를 위해 K마트는 기상요인을 포함한 여러 가지 요인으로 제품을 구분해 제품 코드를 붙이고 있다. 예를 들면, 제품을 세면 도구류, 잡화류 등으로 분류하기보다는 '기상 변화가 심하고 소득이 높은 노년층이 모여 사는 지역에서 초여름에 많이 팔리는 제품군' 등으로 분류하는 것이다.

일본의 경우 전국적인 기상예보는 기상청에서 하고 있지만, 민간 사업자의 기상예보 사업에의 참여를 허가하여 실제로 많은 민간회사들이 기상정보를 제공하는 서비스를 수행하고 있다. 경마장, 건축현장 등 다양한 고객층의 요구에 맞춰 필요한 기상 정보를 적시에 제공해주고 있다. 예를 들면, 일본의 편의점에서 이를 적극 활용하고 있다. 그 이유는 편의점에서 판매하는 제품의 70% 정도가 신선도가 중요한 식품류인데, 식품의 매출이 기상의 변화에 매우 민감하게 반응하기 때문이다.

2.3 군집 분석 알고리즘

군집 분석은 같은 군집 안의 객체끼리는 높은 유사성을 갖게 하고 다른 군집들의 객체와는 높은 상이성을 갖도록 군집이라는 클래스로 데이터를 그룹화시키는 과정이다. 상이성은 객체를 표현하는 속성 값에 기초하여 매겨지는데, 보통 거리척도가 사용된다. 이러한 군집 분석은 데이터 마이닝, 통계학, 생물학, 기계 학습 분야 등에 널리 사용되고 있으며, 매우 다양한 군집 알고리즘이 제안되어 왔다. 군집 알고리즘을 선택하는 기준은 사용하는 데이터와 응용프로그램의 목적에 따라 달라진다. 다음은 주요한 군집 분석 방법의 분류이다.

2.3.1 분할방법(partitioning method)

n개의 객체로 이루어져 있는 데이터베이스를 입력

하여 $k(k \leq n)$ 개의 분할된 데이터의 클러스터를 얻으면서 다음 두 가지 사항들을 만족하여야 한다. 첫째, 각각의 분할은 적어도 하나 이상의 데이터 객체를 포함하고 있어야 하며 둘째, 하나의 객체는 정확히 하나의 클러스터에만 속해야 한다. k개의 초기 분할로 구분된 클러스터들은 재할당을 거듭하여 클러스터를 향상시킨다. 여기서 클러스터가 향상된다는 것은 동일한 클러스터에 속한 데이터 객체들 사이의 거리는 더욱 가까워지고 상이한 클러스터에 속한 데이터와는 객체가 더 멀어져서 클러스터들마다 뚜렷한 차이가 날수록 더 좋은 군집이라 할 수 있다.

2.3.2 계층적 방법(Hierarchical method)

계층적 방법은 계층을 나누는 방식에 따라 크게 집괴식(agglomerative), 분리식(divisive) 두 가지로 나눌 수 있다. bottom-up 접근 방식으로 불리기도 하는 집괴식 방법은 데이터 객체를 분리된 집합으로 나누면서 시작한다. 이후, 반복적으로 유사한 집합과 병합 과정을 거치면서 집합이 가장 최상위 계층에 이를 때까지 이를 반복한다. Top-down 접근 방식이라 불리는 분리식 방법은 모든 데이터 객체가 하나의 클러스터에서부터 시작한다. 이후, 반복적인 분할 과정을 통하여 클러스터가 더 작은 하부 클러스터로 나뉘지게 된다. 이러한 분리 과정은 하나의 클러스터에 하나의 데이터 객체가 할당될 때까지 반복되거나 혹은 종료 조건에 도달될 때까지 반복된다.

2.3.3 밀도 기반 방법(Density-based method)

대부분의 분할방식 군집 방법은 두 데이터 객체 사이의 거리를 기반으로 클러스터를 구성하기 때문에 구(spherical) 형태의 클러스터만을 찾을 수 있다. 때문에 임의의 모양을 하고 있는 클러스터는 찾기 어렵게 된다. 밀도 기반의 군집 방법은 이웃의 밀도가 기준치 이상 되면 클러스터를 확장시켜 나가는 방법이다. 이러한 방법은 상용하면 임의의 모양을 갖춘 클러스터를 찾아 낼 수 있으며 잡음(noise) 또한 제거할 수 있게 된다.

2.3.4 격자 기반 방법(Grid-based method)

격자 기반 방식은 객체 공간을 격자 구조의 유한한 칸으로 양자화 한다. 모든 군집 분석은 이러한 격자 구조에서 행해진다. 이러한 방식의 가장 큰 장점은 빠른 속도에 있다. 격자 기반 방식 군집은 데이터 객체의 수에는 독립적이며 단지 격자 칸(cell) 수에만 영향을 받게 된다.

3. 기후 데이터 모델

시간에 따라 변화하는 기후 데이터의 시공간 특성을 고려한 데이터의 저장 및 체계적인 관리를 위해 관계형 데이터 모델을 적용한다[10]. 또한 기후 데이터베이스는 시간지원 데이터베이스에서 지원하는 시간 속성을 갖는 데이터 관리 기능을 갖추어야 한다. 이 장에서는 기후 데이터를 체계적으로 저장 및 관리하기 위한 기후 데이터베이스의 기본 개념이 되는 시간에 대한 정형화와 기후 데이터의 시간 모델과 데이터베이스의 구조를 설계한다.

3.1 시간의 정형화

시간의 변화에 따라 발생하는 데이터의 이력 정보를 처리할 수 있도록 시간에 대한 명확한 의미(semantics)를 데이터 모델이 지원할 때 데이터베이스를 이용하는 사용자 입장에서는 간편하게 질의어를 이용하여 시간 개념을 갖는 데이터에 접근하여 효과적으로 활용할 수 있다[11]. 또한 시간지원 연산에서는 처리되는 시간의 의미에 대한 정형화가 필요하며, 시간의 종류에 따른 직교성과 시간 데이터 타입의 특성에 따른 결과의 상태 천이를 고려해야 한다. 이와 같이 시간지원 데이터베이스는 시간의 흐름에 따라 변화하는 객체의 이력을 손실 없이 저장 및 관리하는 데이터베이스로 객체가 실세계에서 발생한 시간을 의미하는 유효시간의 시작과 종료를 통해 객체의 이력 정보를 저장 및 관리한다.

3.2 기후 데이터의 시간 모델

시간은 끊임없이 변화하는 현실세계에서 하나의 필수적인 정보로서 의미를 가지며, 사실이나 데이터는 시간과 관련하여 해석할 수도 있다. 이러한 시간의 구조는 선형(linear), 분기(branching), 주기적(periodic) 시간으로 표현되고, 선형 시간의 경우 시작 시점과 종료 시점으로 정의한다. 또한 시간의 밀도는 시간 선이 정수 형태인 이산(discrete)구조, 유리수 형태인 조밀(dense)구조, 실수 형태인 연속(continuous) 구조로 구분할 수 있다. 아울러 연속적인 시간을 표현하기 위해서는 이산 시점, 시간 간격, 그리고 시간을 집합화하기 위한 요소가 있어야 한다[4].

시간지원 데이터베이스는 기존의 관계형 데이터베이스

를 기반으로 시간차원 속성을 추가함으로써 데이터에 대하여 현재값 뿐만 아니라 과거에 발생했던 값이나 미래에 발생할 수 있는 값에 대한 효율적인 관리 기능을 제공한다. 이 논문에서 지원하는 기후 데이터의 시간 모델은 선형이고 이산적이며, 유효시간의 시작 시점과 종료 시점이 있는 범위를 갖는다.

3.3 기후 데이터베이스 구조

기후 데이터의 체계적인 저장 및 관리를 위해 관계형 데이터 모델을 기반으로 하는 데이터베이스를 설계한다. 속성 이력 릴레이션에는 데이터가 측정된 각 지역을 구분하는 지역 식별자와 유효시간, 그리고 일반 속성 정보를 저장하는 릴레이션 구조로 <표 1>과 같은 형태를 갖는다. <표 1>의 속성 이력 릴레이션 구조에서 rid는 기후 데이터가 측정된 국내의 각 지역을 구분하는 지역 식별자(region identifier)이다. 속성 이력 릴레이션에는 각 지역의 유효 시간 간격별로 변화된 속성 정보가 저장된다. 각각의 튜플에 저장되는 유효 시간의 간격은 [vtstart, vtend]로 표현된다.

<표 1> 속성 이력 릴레이션 구조

rid	ts	te	g_1	...	g_n
지역 식별자	유효 시간 시작 시점	유효 시간 종료 시점	속성 정보_1	...	속성 정보_n
string	date	date	float	...	float

<표 2>는 기후 데이터 릴레이션으로 구성된 데이터베이스 테이블을 예시한 그림이다. 표에서 schema는 ilsa(전일사량), precipitation(강수량), ilcho(일조율), cloud(운량), rate(일사율), clean(청명일수), temp(기온), humi(상대습도), wind(바람)와 관련된 속성 데이터를 저장하는 구조이다.

<표 2> 기후 데이터의 스키마 유형

구분	데이터 유형
schema	ilsa, precipitation, ilcho, cloud, rate, clean, temp, humi, wind

속성 이력 릴레이션은 <표 3>과 같이 구성된다. schema는 전일사량, 강수량, 일조율, 일사율, 청명일수, 운량, 기온, 상대습도, 바람에 대한 속성들의 이력

정보를 저장하며, 유효 시간 주기는 월(month) 단위로 사용된다.

<표 3> 기후 속성 릴레이션 : schema

rid	vtstart	vtend	ilsa	prec	ilcho	rate	cloud	clean	temp	humi	wind
r1	1996/01/01	1996/01/31	1978	16	612	65	36	14	-47	69	13
r2	1996/12/01	1996/12/31	1586	11	541	57	38	12	-18	78	9
r3	1997/01/01	1997/01/31	1851	17	495	63	38	14	-58	75	13
r4	1997/12/01	1997/12/31	1414	38	490	61	42	10	-9	79	12

4. 기후 데이터를 이용한 군집 분석

4.1 기후 데이터의 K-means 군집 기법

이전 장에서 설계된 기후 데이터베이스를 기반으로 하여 국내의 기후 데이터를 지역적 특성에 따라 군집하기 위해 이 논문에서 적용한 군집 분석 방법은 k-means 기법이다. 이 방법은 계산 속도가 빠르고 대량의 자료에서 군집을 발견하는데 상당히 효과적인 것으로 알려져 있으며, 다음과 같은 절차로 처리된다.

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \dots & x_{pn} \end{pmatrix} \quad (\text{식 1})$$

단계 1. p 개의 변수로 기술되는 n 개의 자료에 대해 다음과 같은 자료 행렬을 만든다. 즉, X는 p차원에서의 n개의 벡터들을 나타내다고 볼 수 있다.

단계 2. 벡터의 형태로 k개의 군집의 위치를 임의로 결정한다. 여기에서 군집의 개수 k는 분석자가 결정한다.

단계 3. 각각의 자료에 대해 k개의 평균 벡터와의 유클리드 거리를 계산하고 가장 가까운 군집에 그 자료를 할당한다.

단계 4. 각 군집에 대해 평균 벡터를 다음과 같이 다시 계산한다. 여기서 Nm은 m 번째 군집에 속한 자료의 개수이다.

$$Y_{mj} = \frac{1}{N_m} \sum_{i=1}^{N_m} X_{ij} \quad (\text{식 2})$$

(j = 1, 2, ..., p ; m = 1, 2, ..., k)

단계 5. 각각의 자료에 대해 그 자료가 속한 군집(m)을 포함한 모든 군집의 평균 벡터와의 거리를 계산하고 다음과 같은 조건이 만족되면 그 자료가 속한 군집을 바꾼다. (식 3)는 m 번째 군집에 속한 i 번째 자료가 j 번째 군집에 속해 있다고 가정할 경우 두 군집의 평균 벡터와의 차이를 의미한다.

$$\frac{N_m}{N_m+1} \sum_{i=1}^p (X_{ii} - Y_{mi})^2 - \frac{N_j}{N_j-1} (X_{ii} - Y_{j})^2 > 0 \quad (\text{식 3})$$

단계 6. 자료의 재할당이 없을 때까지 4, 5의 단계를 반복한다. k-means 기법에서 군집의 개수 k는 분석자가 정할 수 있다. 하지만 최종적인 군집의 개수는 통계적으로 결정된다. 이를 이용한 군집 개수의 결정 방법은 먼저 군집 개수를 달리하여 군집 분석을 실시한 후, 이때마다 군집기준의 값을 얻고 군집 개수와 군집기준의 값을 그려보아 국소적인 최고값이 나타나면 여기에 대응되는 군집의 개수를 주어진 자료에 가장 적합한 군집의 개수로 결정한다.

4.2 자료

이 논문에서 사용된 자료는 1982년에서 1998년까지의 한국에너지연구원에서 관측된 기후 데이터와 1995

년부터 2000년까지의 기상청의 강수량 데이터이다. 기 후 데이터의 지역 구분을 위해서는 각 데이터의 특성을 나타낼 수 있는 변수를 선택하여야 한다. 예를 들어 강수 지역 구분을 위해서는 일 강수량 자료로부터 각 관측소의 강수 특성을 나타낼 수 있는 변수를 선택 하여야 하며, 이를 위해 여름 6, 7, 8월의 3달에 걸쳐 각 관측 지점에 대한 총 강수량(mm), 강수일 비율(%), 평균 강수 지속일(일), 평균 무강수 지속일(일)을 구하고 다시 전체 관측 기간에 대해 평균하여 군집 분석의 변수로 사용하였다. <표 4>는 이 변수들에 대한 간단한 통계량을 나타내고 있다.

<표 4> 강수량의 간단한 군집 분석 통계

Variable	Average	Standard deviation	Coefficient of variation	
Total precipitation (mm)	655.52	96.04	14.65	
6월	First 10-day (mm)	46.15	12.23	26.51
	Middle 10-day (mm)	40.52	11.49	28.35
	Last 10-day (mm)	67.94	19.26	28.35
7월	First 10-day (mm)	73.25	20.34	27.77
	Middle 10-day (mm)	101.79	20.10	19.74
	Last 10-day (mm)	78.43	18.64	23.76
8월	First 10-day (mm)	63.06	15.45	24.50
	Middle 10-day (mm)	75.83	18.22	24.03
	Last 10-day (mm)	108.57	16.22	14.94
precipitation intensity (mm/day)	18.15	2.29	12.60	
Rainy day percentage (%)	39.25	3.36	8.56	
Duration of rainy day (day)	3.19	0.16	4.88	
Duration of no-rainy day (day)	4.78	0.37	7.80	
Topography (m)	87.80	125.93	143.42	

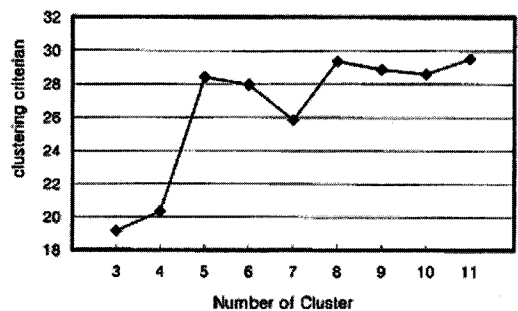
<표 4>는 여름 6, 7, 8월의 3달에 걸쳐 각 관측 지점에 대한 총 강수량(mm), 강수일 비율(%), 평균 강수 지속일(일), 평균 무강수 지속일(일)을 구하고 다시 전체 관측 기간에 대해 평균하여 군집 분석의 변수로 사용 하여, 이 변수들에 대한 간단한 통계량을 나타낸 것이다.

<표 4>에서 볼 수 있듯이 총 강수량의 경우 655.5mm로 연 강수량의 절반가량이 6, 7, 8월에 집중 되어 있음을 알 수 있다. 순별 강수량의 경우 7월 중순 과 8월 하순에 100mm 이상의 강수를 보이고 있으며, 6월 상순과 중순에는 50mm 미만으로 다른 기간에 비 해 상대적으로 적은 양의 강수가 있다. 평균 강수 지 속일과 평균 무강수 지속일은 3.1일과 4.7일이며 다른 변수들에 비해서 관측 지점 사이의 편차가 작은 편이 다. 평균 강수일 비율은 약 40%로 이는 92일 중 37일 가량이며 대략 장마 기간과 일치한다.

군집 분석을 위해 선택된 변수의 단위와 그 규모가 다르므로 특정 변수에 의해 군집이 분석되는 것을 피 하기 위해 각 변수를 정규화 하였다. 따라서 순별 강 수량이 가장 주된 군집 분류의 기준이 된다.

4.3 강수량 기준의 군집 분석 결과

<그림 3>은 각 군집 개수에 대한 군집 기준을 나타 내고 있다. 그림에서 나타난 바와 같이 국소적 최고값 은 군집의 개수가 8개일 경우이다. 또한 군집의 안정성 을 검토하기 위해 주어진 자료에서 임의로 10개의 자료 를 제거하고 군집의 개수를 8개로 하여 다시 군집 분석 을 실시하였으며, 그 결과 각각의 자료는 전체 자료를 사용하여 분석한 경우의 군집에 동일하게 할당되어졌 다. 따라서 주어진 자료에서 얻어지는 군집의 개수는 8 개가 적절하며, 이 군집들이 안정함을 알 수 있다.



<그림 3> 클러스터 수에 대한 군집 기준(강수량)

군집의 개수를 8개로 한 결과를 보면 다음과 같다. 군집 1: 경기도와 강원 영서 지방을 포함하는 지역으로 주로 7월에 많은 강수가 나타난다. 강수 강도가 강하며 강수일 비율을 전체 평균과 비슷하다. 군집 2: 강원 영동 지방으로 8월 상순과 하순을 제외한

모든 기간에 걸쳐 전체 평균보다 작은 강수를 나타낸다. 강수일 비율은 크나 강수 강도는 매우 작은 편으로 소우 지역 중의 하나이다.

군집 3: 전라북도와 충청도 지방으로 7월 중순에 다른 지역보다 많은 강수가 있다. 순별 강수량 이외의 변수로 보면 전체 평균에 가장 가까운 군집이다.

군집 4: 대구를 포함하는 경상도 내륙과 동해안으로 우리나라의 대표적인 소우 지역이다. 전 기간에 걸쳐 다른 지역보다 적은 강수가 내리며, 무강수 지속일 비율이 크고 강수 강도는 작은 값을 갖는다.

군집 5: 남해안 지방으로 7월 상순과 8월 상순을 제외하고는 모든 기간에 걸쳐 다른 지역보다 많은 강수가 내리며 강수 강도 역시 강하다. 하지만 무강수 지속일은 군집 4와 비슷하다.

군집 6: 대관령을 포함하는 태백산맥 산간 지방으로 7월 하순부터 8월에 걸쳐 많은 강수가 내리며 강수일 비율과 평균 강수 지속일이 다른 군집보다 매우 큰 값을 갖는다.

군집 7: 제주도 북부 지방으로 6월 하순을 전후로 한 기간과 8월 중순 이후에 다른 지역보다 많은 강수가 내리며, 강수 강도와 평균 강수 지속일이 크다.

군집 8: 제주도 남부 지방으로 7월 상순까지 많은 강수가 있으며, 강수 강도가 모든 군집 가운데서 가장 크다.

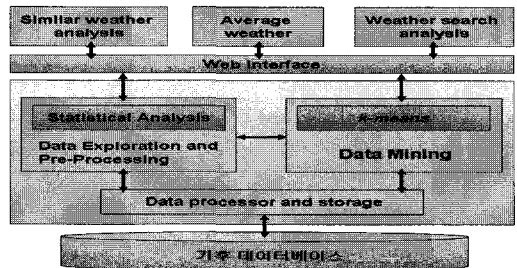
5. 제안시스템의 설계 및 구현

5.1 시스템 구조

이 논문에서 제안하는 기후 데이터 군집 분석 시스템은 <그림 4>와 같으며, 각각의 모듈의 기능은 다음과 같다.

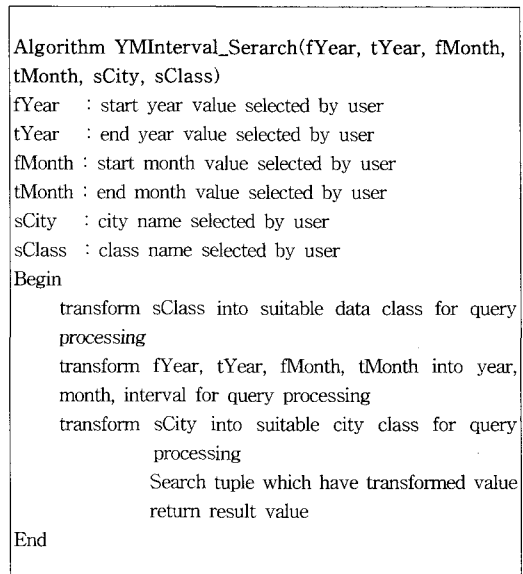
- 데이터 처리기(data processor) : 데이터베이스에 저장된 기후 데이터를 검색하기 위한 데이터 연산을 처리하며, 연산 실행 과정을 거친 후 실행 결과를 반환한다.
- 데이터 마이닝(data mining) : k-means 알고리즘을 적용하여 유사한 데이터들을 유사 지역으로 군집 분석한다.

- 결과 처리(result processing) : 분류된 데이터들을 유사한 지역으로 나타내며 이 데이터들이 날씨 마케팅에 유용한 정보를 제공하도록 지원한다.
- 데이터베이스(database) : 기후 데이터에 대한 정보들이 저장된다.
- 웹 인터페이스(web interface) : 데이터베이스에 저장된 정보를 웹에서 사용자가 직접 검색할 수 있도록 웹 기반의 사용자 인터페이스를 제공한다.



<그림 4> 기후 데이터 군집 분석 시스템 구조

여기서 데이터 처리기는 저장된 기후 데이터의 검색 연산을 수행하는데, 검색은 웹을 통해 저장된 기후 데이터에 대해 지역별, 년별, 월별 검색 등과 같은 이력 데이터 검색을 제공한다. 또한, 검색 결과는 그래프, 차트, 등고선, 월별, 년별 평균 등을 사용자에게 제공한다. 자세한 이력 데이터 검색 과정은 <그림 5>과 같다.



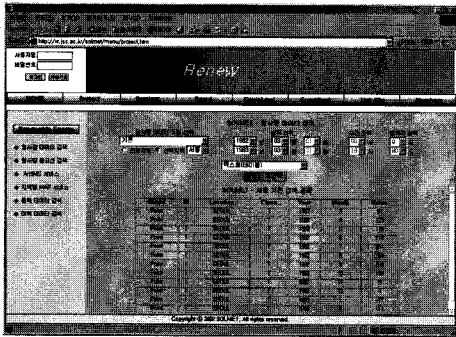
<그림 5> 기후 이력 데이터 검색

5.2 구현 환경

웹 인터페이스 구현 환경은 클라이언트/서버 구조를 기반으로 한다. 시스템의 구현은 Windows 2000 Server(IIS 5.0)를 사용하였으며 데이터베이스는 MS-SQL Server 2000, 웹 프로그래밍 언어는 ASP를 이용하여 시스템을 구현하였다. 웹 인터페이스의 구성 내용은 기후 데이터 검색, 등고선도 검색, 지역별 지도 정보 검색 등의 기능을 제공한다. 검색 결과 유형에서는 테이블(텍스트), 통계 자료, 차트 등의 유형을 선택할 수 있다.

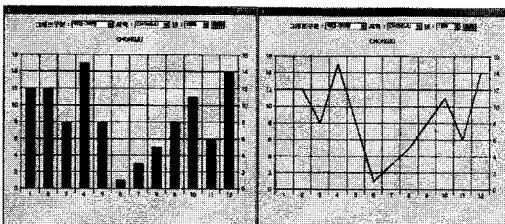
5.2.1 기후 데이터 검색

국내에서 측정된 기후 데이터를 저장 및 관리하기 위하여 웹 인터페이스를 통해 기후 이력 데이터를 관리할 수 있는 기능을 제공한다. <그림 6>은 1982년 3월부터 1983년 3월까지의 서울 지역 기온을 선택한 화면을 테이블 형태로 나타낸 것이다.



<그림 6> 서울 지역 기온 검색 결과 화면 (1982. 03 ~1983. 03)

<그림 7>는 1996년 청주 지역의 청명일수에 대한 막대 1차원, 막대 3차원, 꺾은선 그래프의 검색 결과를 나타낸다.



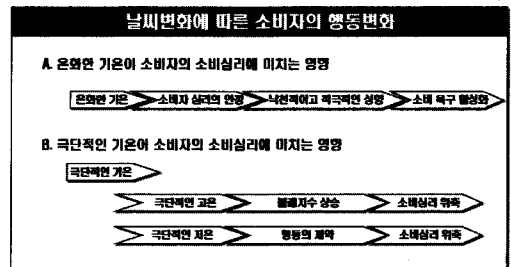
<그림 7> 청주 지역 청명일수의 각종 그래프 검색 결과(1996년)

6. 날씨 마케팅의 응용 사례 연구

이 장에서는 제안된 시스템에 의해 분석된 기후 데이터 정보가 어떻게 날씨 마케팅에 적용될 수 있는가에 대해 기술하기 위해 기후 데이터 정보에 따른 소비자패턴의 변화 및 매출 연관 관계와 날씨 위험관리를 위한 예측 모형 도출 과정, 그리고 그 결과물을 이용하여 백화점에 적용되는 기존의 날씨 마케팅에 대한 연구 사례 [12],[13]를 제시한다. 제안된 시스템은 기후 데이터를 지역적 특성에 따라 군집화 하여 다양한 분석을 가능하게 함으로써 날씨에 기반한 기업의 마케팅 전략을 수립하기 위한 분석 도구로써, 사용될 수 있다.

6.1 날씨에 따른 소비자패턴변화 및 매출 연관 관계

날씨변화에 따른 소비자의 행동 변화는 기업의 마케팅 전략에 매우 중요한 정보이다. <그림 8>은 날씨 변화에 따른 소비자의 행동 변화를 나타낸 것이다.



<그림 8> 날씨 변화에 따른 소비자의 행동변화

<그림 9>은 기온에 대한 데이터를 5도씩 구분하여 상품에 대한 매출에 사용되는 예를 나타낸 것이다. 소프트웨어 0-5도 사이 91%, 맥주/소주/양주 16-20 사이에서 마이너스, 아이스크림 5도 상승시 30%이상의 매출 성장률을 보이고 있다.

상품명	-21~0도	0~5도	5~10도	11~15도	16~20도	21~25도	26~30도	30도 이상
맥주	12	8	18	5	3	-7	18	18
소주	8	8	11	4	-4	-11	8	1
양주	9	8	12	-2	8	-12	11	-8
도시락	-4	8	14	4	7	7	-8	-2
아이스크림	16	8	26	8	18	18	32	32
탕	0	0	-5	3	3	3	-18	2
소프트드링크	4	8	21	6	14	11	11	15

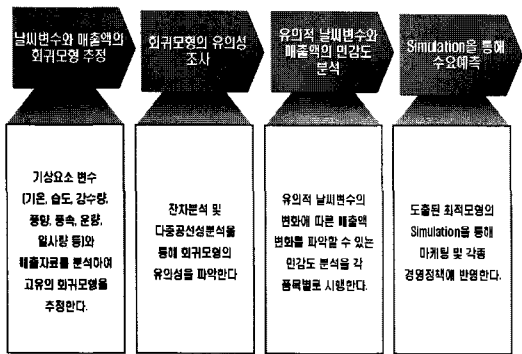
(출처: 산업기상연구소)

<그림 9> 기온 변화 시마다 매출 신장률

이와 같이 기후에 대한 변화 패턴에 따라 제품의 매출 성장률과 밀접한 관계가 있음이 나타났다. 따라서 이와 같은 정보를 자세히 분석 제공하면 기업의 매출 성장에 많은 도움을 줄 수 있을 것으로 판단된다.

6.2 날씨 위험 관리 (WRM : Weather Risk Management)

날씨 위험을 관리하기 위해 필요한 수요 예측 단계는 4단계 과정을 거치게 되는데, <그림 10>은 각각의 단계를 나타내고 있다.

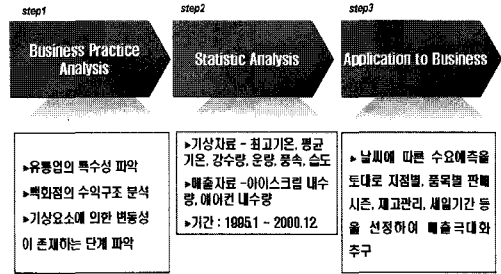


<그림 10> 수요예측 모형 도출 과정

- 1단계: 먼저 매출에 영향을 주는 기후요소를 추정한다. 추정되어진 각각의 기후변수와 매출 자료를 이용하여 회귀모형을 추정하는 단계이다.
- 2단계: 잔차 분석을 통해 회귀검증의 주요 가정의 부합성을 파악하여 회귀 모형의 오차를 줄이는 단계이다.
- 3단계: 독립변수간의 상관관계를 파악하며, 독립변수간의 설명력을 분석하는 단계이다.
- 4단계: 마지막으로 1단계~3단계 분석을 종합한 후, 최적 모형을 도출하는 단계이다.

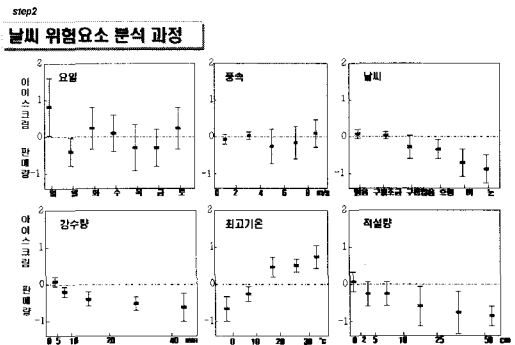
6.3 날씨 마케팅의 적용

수요 예측 모듈에서 나온 결과물이 백화점에 적용되는 과정은 <그림 11>와 같이 3단계를 거치게 된다. 백화점에서 이 자료를 이용하는 목적은 기상요소에 따른 매출변동을 예측하여 효과적인 마케팅전략을 수립하기 위해서이다.



<그림 11> 백화점의 영업 전략을 위한 분석 과정

각각의 과정에서 먼저 1단계는 제품의 구성을 적절히 조절하여 기후변화에 따라 최대의 수익을 낼 수 있는 방법을 모색하는 단계이다. 백화점의 고객은 강우, 강설, 기온 등의 기후요소로 인해 쇼핑여부와 선호매장 등이 다르게 나타나므로 이러한 날씨에 따른 고객수, 단가 등의 변화를 분석함으로써 실질적인 매출의 예측이 가능하고, 주력 상품군을 선정할 수 있다. 예를 들면 여름의 경우 의류, 식료품, 가전제품 등 높은 비중을 차지하는 것을 주력 상품군으로 선정한다. 또한 주요 매출원인 의류, 식료품, 전자제품 등은 특히 날씨의 영향을 많이 받는 제품군으로 그 수요의 정확한 파악을 통한 판매전략 수립이 필수적이다. 이와 같은 조사를 통하여 날씨에 따른 제품별 수요예측을 통한 날씨 마케팅 계획을 수립할 수 있게 된다.

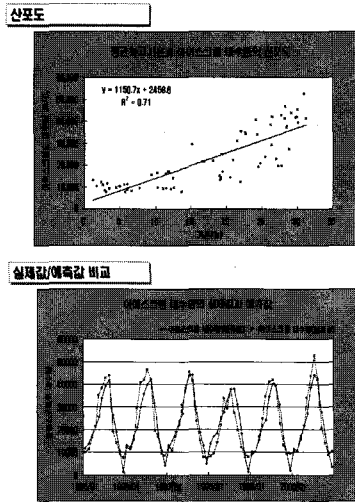


(출처: 산업기상연구소)

<그림 12> 날씨 위험 요소와 아이스크림 판매량과의 분석 과정

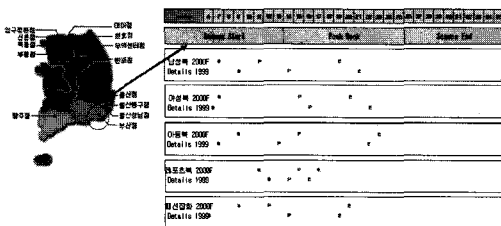
2단계는 통계적 기법을 이용하여 날씨 위험 요소를 분석하는 단계이다. <그림 12>는 날씨 위험 요소에 따른 아이스크림 판매량과의 분석 과정을 예로 나타낸 그림이다. 여기서 데이터는 국내 A백화점의 1995.

1월부터~2000. 1월까지의 5년간 데이터를 적용하였다. 그리고 <그림 13>은 2단계 과정 중 평균최고기온과 아이스크림 내수량의 산포도 및 실제값과 예측값을 비교한 그래프 관계를 나타낸 것이다.



<그림 13> 평균최고기온과 아이스크림 내수량의 산포도 및 실제값과 예측값 비교

마지막 3단계는 위의 1단계와 2단계 분석 결과를 이용하여 총 다섯 가지로 세분화시켜 분석하는 단계이다. 각각의 세분화된 단계는 첫째, 지역별 판매 시즌, 재고관리, 세일 기간 설정 등을 선정하여 효율화 및 매출 극대화를 가져올 수 있도록 한다. 둘째, 과거판매 대비 판매 예상 비율을 조사한다. 셋째, 주단위의 판매 시즌을 예측한다. 넷째, 지점별/제품별 수요를 예측한다. 다섯째, 연간 Promotion Plan으로 연간 계획을 날씨정보에 기초하여 수립 후 매월 Promotion Calendar를 작성하여 정확한 수요예측을 통한 경쟁우위를 확보한다. <그림 14>는 지점별/제품별 시즌의 시작, 절정, 마지막 주로 구분하여 수요의 예측 방법을 나타내고 있다.



<그림 14> 지점별/제품별 주 단위 수요 예측

7. 결론

현대에 이르러 날씨가 기업이나 국가 심지어 개인의 일상생활, 경제 활동에 크나큰 영향을 미친다는 사실은 정설로 받아들여지고 있다. 하지만 기상을 바꿀 수는 없더라도, 향후 비교적 정확한 날씨를 예측할 수 있고, 그것이 기업 활동에 미치는 영향을 정확히 파악할 수만 있다면 기업은 더욱 정확한 수요예측 등이 가능하게 되고, 이를 바탕으로 더욱 효율적인 마케팅 활동을 수행할 수 있을 것이다. 그러나 국내에서는 이와 같은 기상 변화에 의한 소비자 행동 변화에 관한 연구는 아직까지 초보적인 단계이다. 기존에 날씨를 제공하는 기상청 및 관련 업계에서는 단순한 정보를 제공하는 수준에 그치고 있으며 기후 데이터의 특성에 따른 지역성 구분 및 기후 데이터가 어디에 유용하게 사용되는지에 대한 구체적인 정보 제공은 못하고 있다.

이 논문에서는 첫째, 기후데이터를 체계적으로 저장 및 관리할 수 있는 이력 데이터 모델을 제시하고 둘째, 기후 데이터의 특성을 분석한 지역성 구분을 위해 k-means 기법을 사용하였다. 셋째, 측정된 데이터를 웹을 이용하여 데이터베이스에 저장한 후 다양한 연산 및 검색 결과를 제공하였다. 넷째, 기후 데이터를 구체적으로 활용하기 위한 예로서 날씨 마케팅 전략을 도출해내는 과정과 실제 적용을 통해 사전 수요와 피해의 예측을 통해 이익을 극대화 하고 손실을 최소화할 수 있음을 보였다.

제안된 시스템은 농업, 축산업, 수산업, 공업, 건축 에너지, 제조업, 패션, 유통, 관광, 레저, 항공, 환경 등에 효율적으로 활용될 수 있을 것으로 기대한다. 향후 연구될 방향으로는 군집 이외의 다른 데이터 마이닝 기법을 연구하여 좀 더 빠르고 정확한 분석과 날씨 마케팅 및 관련 산업의 활용 방안을 위한 다양한 연구가 필요하다.

참고문헌

[1] L. M.Murphy, J. Brokaw, J. Pulaski, K. McCormack, "The National Alliance of Clean Energy Business Incubators", NREL/BK-720-28724, 2000.

[2] Fsec, "Automated Field Data Management and Quality Assurance", Florida Solar Energy Center, 2003.

[3] K. H. Ryu, "A temporal database management main memory prototype", TempIS TR-26, CSD, The University of Arizona, 1991.

[4] A. Segev, and A. Shoshani, "Logical Modeling of Temporal Data", In Proceedings of the ACM SIGMOD Annual Conference on Management of Data, pp. 454-466, 1987

[5] B. A. Wielicki et al., "CERES(Clouds and the Earth's Radiant Energy System) Validation Plan Overview", Technical Document, Release, 2000

[6] U. Fayyad, D. Haussler, and P. Stolorz, KDD for science data analysis: Issues and examples. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining(KDD-96), 50-56. CA:AAAI Press, 1996.

[7] 오인배, 안윤애, 김원태, 류근호, 김광득, "대체에너지원 이력 데이터 관리 시스템 설계", 한국정보처리학회 논문지, 제10-A권 제6호, 2003

[8] 류근호, 김동호, 남광우, 신예호, 신정훈, "시공간 데이터베이스 시스템 연구", 연구개발결과보고서, 정보통신부, 1998.

[9] 강나영, "시공간 데이터를 위한 클러스터링 기법의 성능 비교", 이화여자대학교 과학기술대학원 컴퓨터학과 이학석사 학위논문, 2003.

[10] 이태규, "국내 일사량 분석 평가 및 데이터 표준화 연구", 산업자원부 최종보고서, 1999.

[11] 김삼남, "시간지원 패러다임에서 개체-관계 모델의 객체지향 모델의 변환", 충북대학교 대학원 전자계산학과, 박사학위 논문, 1997.

[12] 김동식, "기상현상과 산업-날씨마케팅", 케이웨더(주) <http://kweather.or.kr>, 2003.

[13] 이동규, 박정근, "군집분석을 이용한 남한의 여름철 강수 지역 구분", 한국기상학회지, 제35권 제4호, 1999.



이양구

2002년 청주대학교 컴퓨터정보공학과 (공학사)

2004년 충북대학교 대학원 전자계산학과 (이학석사)

2005년~현재 충북대학교 대학원 전자계산학과 박사과정
 관심분야 : 시공간 데이터베이스, 이동객체 데이터베이스, 메인메모리 데이터베이스, LBS, 유비쿼터스 컴퓨팅 등



김원태

2002년 청주대학교 컴퓨터정보공학과 (공학사)

2005년 충북대학교 대학원 전자계산학과 (이학석사)

관심분야 : 시공간 데이터베이스, 지리정보시스템, 지식기반 시스템 등



정영진

2000년 충북대학교 전자계산학과 (이학사)

2002년 충북대학교 대학원 전자계산학과 (이학석사)

2003년~현재 충북대학교 대학원 전자계산학과 박사과정
 관심분야 : 이동 객체 데이터베이스, 이동 객체 색인, Temporal GIS, 유비쿼터스 컴퓨팅 및 질의 처리 등



김광득

1987년 대전산업대학교 전자계산학과 (공학사)

1999년 전북대학교 대학원 전산통계학과 (이학석사)

2000년 충북대학교 대학원 전자계산학과(이학박사)
 1981년~현재 한국에너지기술연구원 책임기술원
 관심분야 : 컴퓨터 보안, 시공간 데이터베이스, 데이터마이닝, 네트워크 관리



류근호

1976년 숭실대학교 전산학과(이학사)

1980년 연세대학교 공학대학원

전산전공(공학석사)

1988년 연세대학교 대학원 전산전공

(공학박사)

1976년~1986년 육군군수 지원사 전산실(ROTC 장교),
한국전자통신연구원(연구원),
한국방송통신대 전산학과(조교수) 근무

1989년~1991년 Univ. of Arizona Research
Staff(TempIS 연구원, Temporal DB)

1986년~현재 충북대학교 전기전자 및 컴퓨터공학부
교수

관심분야 : 시간 데이터베이스, 시공간 데이터베이스,
Temporal GIS, 유비쿼터스 컴퓨팅과 스트림
데이터, 지식기반 정보검색 시스템, 데이
터마이닝, Bio-Informatics