한국수학사학회지 제19권 제3호(2006년 8월), 113-122

평활 방법론이 적용될 수 있는 컴퓨터-소프트웨어 교육분야 제안

서경대학교 소프트웨어학과 이승우
swlee@skuniv.ac.kr

본 논문에서는 오늘날 관심의 대상이 되고 있는 평활에 관한 수학적 배경, 통계적 방법론, 그리고 평활방법론이 적용될 수 있는 컴퓨터-소프트웨어 교육분야에 관하여 조사하고 논하고자 한다. 뿐만 아니라 수학과 통계를 기반으로 히스토그램, 커널밀도추정량, 적응커널추정량, 띠너비선택방법에 관한 개념과 방법론을 소개하고자 한다.

주제어 : 히스토그램, 커널밀도추정량, 적응커널추정량, 띠너비선택방법

0. 서론

전 세계가 지식정보화 사회의 도래로 무한 경쟁시대에 본격적으로 돌입했으나 우리나라는 근본적인 경쟁력 부족의 문제에 봉착해 있다. 우리나라 대학의 컴퓨터-소프트웨어 관련학과는 이미 많은 인력을 배출하고 있으나 컴퓨터-소프트웨어 인력양성 및 공급은 총체적 위기에 처해있다. 그 원인에는 여러 가지가 있겠으나 창의적 아이디어와 능력부족, 문제해결 능력의 결여가 주된 요인들 중에 하나이다.

이러한 문제점을 해결하기 위한 방편으로는 수학 및 통계를 이용하여 기본 원리를 숙지하고 창의적으로 활용할 수 있는 능력, 문제해결 능력을 갖춘 인력을 양성하여야 한다. 지식 정보화 시대를 선도하는 인재 양성은 스스로 문제를 정의하고 그 해를 찾는 과정 중에서 창의력 개발도 가능하다.

오늘날 우리나라의 현실에 비추어 볼 때 국제경쟁력을 확보하기 위해서는 지식정보화 사회의 새로운 환경에 맞는 사고를 배양하는 것이 절대적으로 필요하다. 컴퓨터-소프트웨어교육의 근본적인 한계를 극복하고 내실화하기 위한 현실적인 방편으로, 본 논문에서는 수학 및 통계를 기초로 한 컴퓨터-소프트웨어 교육을 주장한다.

이에 대해 컴퓨터-소프트웨어 교육에서 수학 및 통계를 어떠한 방법으로 가르칠 것인가는 하는 근본적인 문제를 해결해야한다. 즉 컴퓨터-소프트웨어 교육을 위해 수학 및 통계의 기본적인 이해가 필요하다. 수학 및 통계의 이해에는 개념의 발달과정과 관련하여 그 의미를 찾고, 역사-발생적 분석을 통하여 그 의미를 살펴보고 활용하는 것이 효과적이다.

특히 본 논문에서는 수학 및 통계의 평활(smoothing)에 초점을 두고자 한다. 컴퓨터-소프트웨어 교육과정에서 평활의 연계 가능성을 분석하면서 평활에 적용해 본 사례들을 살펴보고 그 가능성을 제안하고자 한다. 또한 컴퓨터-소프트웨어분야에서의 수학 및 통계의 역할과 의무가 막중함을 보이고자한다.

확률밀도추정에서부터 발전된 평활방법론(smoothing technique)은 1857년 경제학자 엥겔(Engel)이 regressogram을 고안한 것을 시초로 오랜 전통을 가지고 있다. 그러나 본격적으로 연구된 것은 20~30년 전부터이며 그때부터 평활의 이론과 방법론이 체계화되어져갔다. 여러 종류의 평활방법론들이 존재하지만 모두 동일하게 관심을 받아온 것은 아니다. 그 중에서 히스토그램, 커널추정량(kernel estimator), 적응커널추정량(adaptive kernel estimator) 그리고 스프라인추정량(spline estimator)이 수학적인 이론을 근거로 발전되어왔으며 오늘날 관심을 받고 있는 평활 방법론이 되었다.

위에서 제시된 방법론이 주목되는 이유는 두 가지가 있다. 첫째, 이 추정량은 구조적으로 간단하고 계산하기가 쉬우며 수학적으로 접근하기가 수월하다. 두 번째, 이 추정량은 응용할 수 있는 분야가 많으며 광범위하게 적용할 수 있다는 장점이 있다.

본 논문에서는 21세기 IT산업 발전을 도모하기 위하여 컴퓨터-소프트웨어분야에 통계적 평활기법을 도입하여 교육시킴으로서 더 생산적이고 바람직한 기술을 창출하고자 연구하였다. 그리고 평활기법이 컴퓨터-소프트웨어의 다양한 분야에서 어떻게 응용되는지를 살펴보고자 한다. 1장에서는 오늘날 IT분야에서 중요하게 사용되고 있는 평활기법인 히스토그램을 적응커널추정량(adaptive kernel estimator)으로 대치시키기 위하여 히스토그램이 적응커널추정량으로 발전해가는 과정을 살펴보고자 한다. 2장에서는 IT분야에서 사용되고 있는 히스토그램의 평활기법을 적응커널추정량으로 적용하고자 한다. 그리고 IT분야에 평활기법을 상호보완적으로 연관시켜서 단계적으로 활용될 수 있도록 적용한 예와 적용할 수 있는 교육분야에 관하여 분석하고 제안하고자 한다. 3장에서는 결론을 맺는다.

1. 평활방법론(Smoothing Technique)

표본의 크기가 증가하여 무한대일 경우, 히스토그램의 구간의 간격이 좁아져서 가상적인 곡선에 근사함을 알 수 있다. 이와 같이 표본의 크기가 무한히 클 때 히스토그램의 극한곡선을 밀도함수(density function)라고 한다.

확률변수 X의 분포를 묘사할 수 있는 기본적인 방법론으로서 확률변수에 관한 확률밀도함수를 사용할 수 있다. 확률밀도함수는 여러 분야에서 유익하게 활용할 수 있는 다양한 정보를 제공한다.

확률변수 X의 분포가 비대칭이거나 다양한 형태로 분포되어 있는 곡선(curve)인

경우, X의 구조적인 요소와 모양으로 확률밀도함수를 해석할 수 있다. 미지의 확률밀도함수 f의 추정은 곡선(curve)에 관한 정보의 이해와 해석을 제공한다. 대부분 실질적인 연구에서 X의 확률밀도함수를 정확히 발견하는 것은 불가능하다. 그러므로 관측값들을 이용한 미지의 확률밀도함수 추정은 지난 수십 년 동안 다양한 방법론으로서 지속적으로 연구 개발되고 있다. 본 장에서는 현재 IT분야에서 평활방법으로 사용되고 있는 히스토그램을 적응커널추정량으로 대치시키기 위해서 히스토그램에서 적응커널추정량으로 단계적으로 변화하는 과정을 통하여 통계학적인 방법론을 알아보고 장단점을 파악하고자 한다.

히스토그램은 미지의 확률밀도함수 f를 추정하기 위해서 최초로 사용된 밀도추정량이며 도수분포를 나타내는 그래프로서 관측값의 분포의 특징에 관한 정보를 시각적으로 제공하는데 사용된다. n개의 관측값이 $X_1, X_{2,} \cdots, X_n$일 때, 히스토그램의 확률밀도함수는

$$\hat{f}(x) = (nh)^{-1} \sum_{i=1}^{n} \sum_{j} I(X_i \in B_j) I(x \in B_j)$$

이다. 여기서 $B_j := [x_0 + (j-1)h,\ x_0 + jh),\ j \in Z$이고 x_0는 히스토그램의 원점이다. $h > 0$는 빈폭(binwidth)이라고 한다. 히스토그램의 통계량은 원점의 선택에 의존하는 단점이 있다. 이러한 단점의 해결책으로서 Silverman(1986)이 방법론을 제시했으며 Scott(1985)는 평균이동히스토그램(average shifted histogram)을 이용하여 평균을 이동함으로서 원점에 의존하지 않음을 보였다.

단순추정량(naive estimator)은 Fix와 Hodges(1951, [1, 32])에 의해 소개되었으며

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} w\left(\frac{x - X_i}{h}\right), \quad x \in R$$

로 정의된다. 여기서 w는 가중함수(weight function)이므로 $X_i \pm h$에서 계단함수(step function)의 속성을 가지고 있으므로 완벽하게 자료의 성질을 표현하지 못하는 경향이 있으며 $\hat{f}$는 미분가능하지 않다.

커널밀도추정(Kernel Density Estimation)은 단순추정량의 단점을 해결하기 위해서 Rosenblatt가 1956년에 소개하였다([1, 32]). 커널함수 K를 이용하여 단순추정량의 가중함수 w로 대치시킴으로서 단순추정량의 단점을 극복했다. 그러므로 커널밀도추정은 관측값들의 형태를 곡선으로 적절히 묘사할 수 있으며 수리적으로 다루기 쉽다는 장점이 있다. 커널밀도추정량은

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

이고 평활의 정도를 조절하는 평활모수(smoothing parameter)인 띠너비(bandwidth) h와 커널밀도함수 K로 이루어진 두 개의 모수(parameter)로 정의된다. 그리고 커널함수 K는 0에 관하여 대칭이고 적분값은 1이며 n번 미분 가능하므로 커널밀도함수도 미분가능하다. 띠너비 h와 커널밀도 K가 유일하다면, 주어진 관측값들에서 이에 상응하는 커널밀도함수는 유일하다. 띠너비를 임의로 결정함으로서 곡선의 형태가 상이해질 수 있으며 이상점이 내포된 확률분포의 자료에 커널추정량을 적용할 때 띠너비를 전제의 표본 집합에 적용함으로서 잡음에 영향을 받을 수 있는 단점이 있다

최근접이웃방법(k-Nearest Neighbor; k-NN)은 Loftsgaarden과 Quesenberry (1965, [1, 33])에 의해서 소개되었다. k번째 최근접이웃 밀도추정량은

$$\hat{f}(x)=\frac{1}{nd_{k(x)}}\sum_{i=1}^{n}K\left(\frac{x-X_i}{d_{k(x)}}\right)$$

로 정의된다. 여기서 띠너비 $d_k(x)$는 임의의 자료 x로부터 k번째 표본의 자료까지의 거리이며 전체 곡선의 평활의 정도는 정수 k의 선택에 의해서 결정된다. k번째 최근접이웃 밀도추정량은 함수 $d_k(x)$가 불연속인 곳에서 미분 불가능하다는 단점을 가지고 있다. k-NN 추정량은 임의의 x로부터 h 근방(neighborhood)을 적절하게 변화시키면서 그 구간에 포함된 관측값들의 국소평균(local average)을 기초로 한다.

가변커널방법(Variable kernel method)은 Breiman, Meisel과 Purcell(1977)에 의해 제시되었고 Wertz(1978)가 연구해서 체계화되었다([1, 33]). 가변커널추정량은 고전적 커널추정과 유사하게 고안되었으나 최근접이웃방법과 관계가 있다.

$d_{j,k}$를 X_j에서부터 k번째로 가까운 자료들간의 거리라고 하면 가변커널추정량은

$$\hat{f}(x)=\frac{1}{n}\sum_{j=1}^{n}\frac{1}{hd_{j,k}}K\left(\frac{x-X_j}{hd_{j,k}}\right)$$

로 정의된다. 최근접이웃추정에서 사용된 띠너비는 x로부터 자료들간의 거리에 종속되지만 가변커널추정에서 사용된 띠너비는 x에 독립이며 자료들간의 거리에만 종속된다.

적응커널추정량(adaptive kernel estimator)은 가변커널 추정량으로부터 유도된다. 확률밀도함수 f를 추정하는데 관심을 가지고 있는 Abramson(1982, [1, 101])은 자료에 적응 가능한 윈도우너비(data-adaptive window width)를 계산할 수 있는 적응커널추정량(adaptive kernel estimator)을

$$\hat{f}(x)=\frac{1}{n}\sum_{j=1}^{n}\frac{1}{h_j}K\left(\frac{x-X_j}{h_j}\right)$$

로 정의했으며 h_j는 평활모수이며 국소 띠너비(local bandwidth)라 한다. 즉, 적응커널추정량은 주어진 구간에서 관측값들에 의해서 자유롭게 띠너비를 조절함으로서 추

정할 수 있는 방법이다. 그리고 위에서 제시한 확률밀도추정량은 불편성(unbiasedness)과 일치성(consistency)을 만족해야 한다. 확률밀도함수 f의 추정량 $\hat{f}$가 모든 $x \in R$에 대하여 $E_f[\hat{f}(x)] = f(x)$ 라면 f에 대한 불편 추정량(unbiased estimator)이다. 그리고 f의 평균제곱오차(mean squared error)는

$$MSE(x) = E_f[\hat{f}(x) - f(x)]^2$$

이고 누적평균제곱오차(integrated mean squared error)는

$$IMSE = \int_{-\infty}^{\infty} E_f[\hat{f}(x) - f(x)]^2 dx$$

이며 적분제곱오차(integrated squared error)를

$$ISE = \int_{-\infty}^{\infty} [\hat{f}(x) - f(x)]^2 dx$$

로 정의해서 일치성의 측도로 활용하여 최적의 띠너비를 구할 수 있다.

커널밀도추정은 밀도 함수 추정에서 커널을 유용한 도구로 사용하고 있다. 커널에서의 띠너비 h는 모수이므로 h의 선택은 커널추정에서 중요한 문제이다. 주어진 관측값들로부터 임의의 곡선을 추정하기 위하여 커널추정을 이용한다면, 어떤 띠너비를 선택하느냐에 따라 추정된 곡선의 표현과 해석은 매우 주관적이다.

만약 h를 감소시키면, 즉 $h \to 0$이면 각 띠너비에 소량의 관측값들이 포함되어서 커널밀도추정을 통하여 매끄럽지 않은 곡선의 형태가 된다. 만약 h를 증가시키면, 더 많은 관측값들이 각 띠너비에 포함되기 때문에 곡선 형태는 보다 더 매끄럽게 된다. 만약 $h \to \infty$이면 곡선은 매우 완만하고 평평한 형태를 가진다. 그러므로 실제적인 비모수 커널밀도 추정하에서 여러 종류의 커널함수을 변경해서 커널밀도추정을 하여도 구해진 곡선의 형태는 별 차이가 없으므로 띠너비 h의 선택 방법이 추정에서 중요한 문제이다.

위에서 제시한 적응커널추정량의 기법은 밀도추정에서 뾰족한 끝, 첨단에서는 과소추정(underestimate)하고 움푹 패어 들어간 곳, 골짜기에서는 과대추정(overestimate)하는 경향을 보임으로써 커널밀도 추정의 또 다른 단점을 제거 또는 감소하는데 적절하다. 그리고 주어진 구간에서 고정된 전체적 띠너비(globally bandwidth)가 아닌 주어진 구간에서 자유롭게 띠너비를 조절할 수 있는 국소 띠너비(local bandwidth)를 사용함으로서 추정의 편의를 제거 또는 감소시킬 수 있는 장점이 있다. 그러나 주어진 구간의 경계근방에서 경계(boundary)의 효과를 제거 또는 감소시켜서야 한다.

관측값들로부터 주관에 의해 구해진 띠너비에 따라 추정값에 의해서 구해진 곡선의 형태는 각기 다른 정보를 제공하기 때문에 곡선을 추정하기 위한 주어진 띠너비가 적절한가를 확인하는 것이 필수적이다. 그러므로 최적의 띠너비는 *MSE*, *IMSE*, *ISE*

를 최소화하는 해를 구함으로써 얻을 수 있다.

그러나 이 띠너비에는 미지의 함수를 포함하고 있기 때문에 h를 합리적으로 찾는 방법론들 중에서 플러그인방법(plug-in method)과 교차검정과정(cross-validation; CV)이 있다. 교차검정과정은 최대우도(maximum likelihood) CV와 최소제곱(least-square) CV가 있다.

위에서 제시한 평활방법론 외에도 추정회귀곡선에 대한 실제 관측값들의 일치성에 관한 측도로서 잔차제곱합을 이용하는 **스프라인 평활(Spline Smoothing)**, 여러 가지 밀도추정방법과는 다른 견지에서 접근하는 방법론으로서 푸리에 전개(Fourier expansion)의 개념을 이용하는 **직교열추정량(Orthogonal series estimator)**, 최대우도(maximum likelihood)와 같은 통계적 방법론을 적용하는 **최대벌점우도추정량 (Maximum penalized likelihood estimator)** 등이 있다.

2. 수학 및 통계를 이용한 컴퓨터-소프트웨어 교육

2.1 컴퓨터-소프트웨어 교육현황과 수학 및 통계의 역할

우리나라 컴퓨터-소프트웨어학과의 졸업생들의 능력이 정보화 사회가 요구하는 기준에는 미치지 못하고 있다는 것이 일반적인 평가이다([3], [4]). 즉 실무적응 능력 및 함양부족, 창의력 부족 등이 그것이다. 그 원인에는 여러 가지가 있겠으나 교과과정의 문제점을 들 수 있다. 최근에 우수한 소수의 대학을 제외하고는 전공학점의 하향조정으로 쉬운 전공과목을 선택하고 수학 및 통계를 기피하는 현상이 나타나고 있다. 더욱이 몇몇 IT분야의 전문가조차도 수학 및 통계의 가치를 인정하지 않고 있는 실정이다([3]). 특히 수학 및 통계는 IT분야에서 기초학문으로만 인정하고 있다.

그러나 수학 및 통계는 IT분야에서의 전공교과목과 아주 밀접한 관련이 있으며 아주 유용하고 중요한 토대로서 작용하고 있다. 수학 및 통계와 관련된 IT분야의 전공기초교과목으로서 대학수학, 통계학, 선형대수, 이산수학 등이 있다. IT분야의 전공 교과목 중 프로그래밍언어, 자료구조, 알고리즘, 수치해석은 수학의 사고력·개념과 밀접한 관련이 있는 교과목이다. 그리고 IT분야의 게임분야, 데이터베이스분야, 인공지능분야는 수학 및 통계가 필수요건이다. 그러므로 수학 및 통계는 인력양성을 위한 IT분야 전공필수교과목으로 지정해야한다. 이제 컴퓨터-소프트웨어학과에서 수학 및 통계의 중요성은 뒤로 미룰 수 없는 시급한 문제가 됐다.

IT분야에서 수학 및 통계가 필요한 두 가지 이유가 있다. 첫 번째 이유로서, 수학은 모든 학문의 기본이기 때문에 그 원리를 이해하고, 개념을 바탕으로 사고력을 키우며, 문제해결능력과 창의력을 배양할 수 있다. 창의적인 사고란 스스로 아이디어를 고안하고 새로운 기술을 발전시켜서 문제를 해결하는 능력이다. 두 번째 이유로서, 수학 및 통계학의 전문지식이 IT분야의 학문적 교류를 통해 원천기술과 기술혁신을 확보한

후 새로운 기술을 창출하는 것이다. 급속히 변화하는 정보화사회에서 국제 경쟁력을 확보하기 위한 방편으로서, 산업현장에서 기존 기술을 단순히 적용하는 수준을 넘어 새로운 기술을 연구하고 창출함으로서 현재 직면한 문제를 상당수준 해결할 수 있다. 2.2절에서는 위에서 제시한 두 번째 이유, 즉 수학 및 통계의 전공지식인 평활을 이용하여 IT분야에 접목해 보고자한다.

2.2 평활 방법론이 적용될 수 있는 컴퓨터-소프트웨어 분야 제안

세계 기술 경쟁에서 첨단 신기술의 창출과 주도권 선점은 지속적인 기초연구와 역량강화로 이루어진다. 그러므로 수학과 통계학이 일상생활에 다양하게 응용되면서 현대 과학기술의 근간이 되어야만 한다. 뿐만 아니라 수학과 통계학이 IT 및 컴퓨터기술의 핵심적인 학문으로서 다양한 기법을 제공해야만 한다.

본 절에서는 현재 IT분야에서 많이 활용되고 있는 히스토그램의 평활기법을 적응커널추정량으로 적용하고자 한다. 또한 히스토그램의 평활기법보다 적응커널추정량을 다양하게 접목시킴으로써 IT산업의 발전을 도모하고자 한다.

오늘날의 컴퓨터는 과거 계산장치에서는 상상할 수 없을 만큼 빠른 속도로 사칙연산과 같은 기본산술뿐만 아니라 각종 수학, 물리학, 천문학 등에서 사용하는 고난도의 수식을 정확하게 계산할 수 있다. 이러한 컴퓨터의 계산기능을 이용하여 판단기능과 자료처리기능에 평활기법을 적용하고자 한다.

평활방법론을 컴퓨터의 판단기능에 적용하면 다음과 같다. 현재 활발히 응용되고 있는 히스토그램의 평활기법 기술은 다음과 같다. 영상의 영역분리기법, 카메라 입력영상의 생동폭 확장기법, 음성인식시스템의 환경잡음 처리기법에 적용되어서 실생활에서 사용되고 있다. 1장에서 제시한 히스토그램의 평활기법에 적응커널추정량을 적용함으로써, 보다 더 많은 잡음을 제거할 수 있기 때문에 영상 및 음성데이터 처리 등의 분야에서 보다 더 정확하고 신뢰성 있게 적용할 수 있다.

또한 일정한 규칙을 갖기 힘든 인식에 활용될 수 있는 인공지능의 전문가시스템(expert system), 신경망(neural networks), 유전자 알고리즘(genetic algorithms)에 적용할 수 있다. 위에서 제시한 인공지능분야는 일정한 시간 안에 혹은 수용할 만한 해가 나올 때까지 테스트한다. 즉 평활기법을 사용하여 유용한 패턴이나 규칙(지식)을 추출함으로써, 컴퓨터의 판단기능을 인간의 판단기능과 유사한 형태로 발전시킬 수 있다.

IT분야 외에서는, 금융기관에서의 경제성장률과 주식의 향후 예측, 일기예보, 도로의 교통량을 예측하여 적절하게 교통신호를 제어하는 등 일상생활에서 향후 발생할 결과를 미리 예측할 수 있는 분야에서 현재 사용되고 있다.

평활방법론을 컴퓨터의 자료처리기능에 적용하면 다음과 같다. 자료[1]는 정리되고

1) 자료(data)는 현실세계로부터 관찰이나 측정을 통해서 문자, 수치, 기호, 음성 또는 그림 등의 형태로 표현한 기록 즉, 사실(fact) 또는 값(value)으로 정의된다.

유용한 형태를 가져야만 정보[2]로써 가치가 있기 때문에 분류(classifying), 정렬(sorting), 계산(calculating), 요약(summarizing)의 단계를 거쳐 검색(retrieval)을 통하여 필요로 하는 자료를 쉽게 찾을 수 있다. 정보관리 기술은 수학과 통계학의 기본개념을 토대로 발전되어왔다. 그러므로 정보산업의 기본이 되는 데이터베이스의 관리 기술도 수학과 통계학의 기본 개념을 토대로 항상 최신의 정보를 제공하기 위해서 발전되고 있다. 즉, 현실세계에서 얻어진 데이터는 어떠한 유용한 처리과정을 통하여 불확실성을 감소시켜주는 정보로 생성하는 과정을 자료처리(data processing) 또는 정보처리(information processing)라고 한다.

현재 자료처리기능에 응용되고 있는 기술로서 데이터베이스와 데이터마이닝[3](datamining)이 있다. 정보처리분야에 통계학의 평활기법을 적용함으로서, 대량의 데이터베이스를 탐색, 분석하여 데이터에 내재하는 유용한 패턴을 발견함으로써 최적의 정보 선택이 가능하다.

세부 분야별(시스템소프트웨어분야, 인터넷응용분야, 멀티미디어분야, 컴퓨터네트워크분야, 게임소프트웨어분야, 지능형 소프트웨어분야, 정보검색 및 검색분야, 기타분야)로 평활 기법이 적용될 수 있는 소프트웨어 분야를 정리하면 다음과 같다.

- **시스템 소프트웨어 분야:** 데이터베이스 관리시스템(DBMS) 설계에서 자료를 통한 최적의 정보 선택, 컴퓨터네트워크의 근거리 통신망에 적용, 컴파일러와 운영체제는 현재로서는 적용 불가능
- **인터넷응용 분야:** 무선 인터넷의 프로그램을 개발하기 위한 제반 기술에 응용, 인터넷방송에 응용
- **멀티미디어 분야:** 디지털 영상처리의 화면 인식에 응용, 멀티미디어 시스템의 영상회의, 컴퓨터그래픽스의 이미지 합성, 멀티미디어 저작에서의 음성분야, 멀티미디어 신호처리에 적용
- **컴퓨터네트워크 분야:** 데이터통신의 기호간의 간섭에 적용, 네트워크프로그램에서의 에러 검색, 분산처리에서의 에러복구에 응용
- **게임 소프트웨어 분야:** 디지털신호처리에 적용, 컴퓨터그래픽스・컴퓨터애니메이션・가상현실에서의 컴퓨터 영상에 적용
- **지능형 소프트웨어 분야:** 인공지능의 지식의 검색・추론・학습 등에 이용, 컴퓨터비전의 시각적 정보를 처리하는데 적용, 인공지능 프로그램은 현재로서는 불가능
- **정보처리 및 검색분야:** 정보검색에서 적용, 정보보호에서 응용, 4세대언어에서의 미래지향적 프로그래밍 기법에 적용 불가능
- **기타 분야:** 소프트웨어 개발 및 유지 보수(Software Creation and Maintenance)에

2) 정보(information)는 의사결정이나 업무수행에 유용하게 활용할 수 있는 지식(knowledge)으로 정의된다.
3) 데이터마이닝은 통신데이터의 다차원분석, 금융 데이터의 분석, 생물의약과 DNA 자료분석 등에서 평활기법을 적용할 수 있다.

적용, 검증(Verification and Validation)에 응용, 전자상거래 시스템 개발 및 인터넷 서비스에 응용

3. 결론 및 제언

지금까지 통계학의 평활기법이 적용될 수 있는 IT분야에 관하여 살펴보았다. 21세기 IT산업은 국가의 경쟁력 제고 차원에서 중요한 역할이 기대되고 있는 분야이다. 우리나라는 천연자원이 부족하지만 인적자원이 풍부하기 때문에 IT산업은 향후 모든 산업 분야에 미치는 영향력이 크고 잠재력을 가진 산업이다. 그러나 현재 우리나라의 IT산업은 총체적 위기에 처해있다. 그 대안으로서 IT사업을 양 중심에서 질 중심으로 전환하여야 한다.

오늘날 순수 이론적인 발전만으로 가치가 인정될 수 없다. 이론과 그의 적절한 응용기법이 병행해서 발전되어야 학문의 목적을 달성할 수 있다. 즉, 수학 및 통계이론과 방법론은 물론이고 그 적용에 이르기까지 연결되어야 그 가치와 목표가 성취되는 것이다. 그러므로 수학, 통계학이 응용 분야와 상호교류를 통하여 발전을 도모할 수 있어야 한다.

본 논문에서는 통계학의 평활기법을 IT분야에 접목시킴으로서 연구교류를 확대 촉진하여 앞으로의 통계학 발전 및 IT분야에 기초적 터전을 마련하고자 한다. 그리고 원천기술의 확보와 기술혁신의 원동력이 되기 위해서는 수학과 통계학이 IT분야와 서로 결합하는 방식이 구체적으로 제시되어야 한다. IT분야는 다른 학문분야보다 새로운 응용의 창출이 요구되는 분야이다. 새로운 기술로 발전시키는 창의적인 사고와 타분야와 공동작업수행능력을 겸비할 수 있는 능력이 배양해야 발전 할 수 있다. 그 방편으로써 컴퓨터-소프트웨어와 수학 및 통계학의 학문적 교류를 통하여 내실을 기해야한다.

본 논문에서는 현실적으로 모집단에 대한 가정이 충족된다고 볼 수 없는 상황이거나 특별한 가정을 필요로 하지 않은 경우가 대부분이므로 미지의 확률밀도함수를 추정하기 위해서 비모수적 접근방법을 사용하였다. 또한 평활방법론 중에서 IT분야에서 많이 사용되고 있는 히스토그램의 대안으로서 적응커널추정량을 적응시켜보았다. 그러므로 수학과 통계학이 이론적인 중요성만을 중시하던 과거와는 달리 여러 다양한 분야에 적용되고 더 실질적인 연구를 진행함으로써 학문이 발전할 수 있다.

본 논문에서는 IT분야에 평활기법을 적용함으로써 컴퓨터-소프트웨어 개발을 위한 방법론을 제안하였으며, 그에 대한 구체적이고 지속적인 연구가 필요하며 실제 적용을 통해서 보완 및 개선이 필요하다고 하겠다.

참고 문헌

1. Silverman, B. W., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York, 1986.
2. 김진형, IT 인력양성의 현황과 개선 방향, 정보처리학회지 제10권 제5호(2003), pp. 8-12.
3. 임순범, 송종수, 엄영익, 산업계 수요에 기반한 컴퓨터-소프트웨어 표준교과과정 분석, 정보처리학회지 제10권 제5호(2003), pp. 34-42.
4. 정책연구 02-15 소프트웨어산업진흥 5개년계획(2003~2007), 소프트웨어진흥원, 2003. 1

On Teaching of Computer-Software Field Using Smoothing Methodology

Department of Software, Seokyeong University **Seung-Woo Lee**

We investigate the mathematical background, statistical methodology, and the teaching of computer-software field using smoothing methodology in this paper. Also we investigate conception and methodology of histogram, kernel density estimator, adaptive kernel estimator, bandwidth selection based on mathematics and statistics.

Key words: histogram, kernel density estimation, adaptive kernel estimator bandwidth selection

2000 Mathematics Subject Classification: 97-03,

ZDM Subject Classification: K85

논문 접수: 2006년 4월 심사 완료: 2006년 6월