



濁流清論

제55호 2018년 12월 13일(목)

발행인: 정영기 / 편집: 편집위원회

< 알려 드립니다 >

◇ 수시로 원고 접수합니다

탁류청론은 아주대학교 교수님들의 의견을 나누는 자리입니다.

교수님들의 원고는 교수회 이메일 (jy717@ajou.ac.kr)로 보내주시시오. 탁류청론에서 다루었으면 하는 주제가 있으면 위의 메일 주소로 보내주시기 바랍니다. 감사합니다.

목 차

<u>더불어 푸른 솔밭에서</u>	1
빅데이터 수학에 대한 단상	
<u>특집</u>	2
1. 임베디드 시스템을 위한 빅데이터 분석	
-이젠 거꾸로! 기계가 하는 말에 귀 기울여야 한다.	
2. 생물정보학 및 유전체학 최신동향	
3. 디지털역사학의 시작과 도전	
4. 의학과 데이터과학의 만남	
<u>소통과 담론</u>	9
교원 임면권 관련 아주대학교 정관 변경에 대한 우리의 우려	
<u>교수들의 건강칼럼</u>	10
우리병원 남자들 관찰기	
<u>소식</u>	12
교수회 제91,92,93차 월례 대의원회 개최	

더불어 푸른 솔밭에서

빅데이터 수학에 대한 단상

데이터사이언스학과 정재훈

뉴욕 대학교의 Courant Institute of Mathematical Sciences에는 그 이름만으로도 별처럼 빛나는 유명한 수학자들이 여럿 있는데 그 중 한 분이 Peter Lax 교수이다. 헝가리 출신으로 편미분방정식(Partial Differential Equations) 분야에서 뛰어난 업적을 낸 수학자인 그는 2005년 아벨상(Abel Prize) 수상을 비롯해서 그 외의 많은 상과 명성을 얻은 우리 동시대의 위대한 수학자 중의 한 분이다. 필자의 지도교수가 Lax 교수의 친한 친구였던 탓에 Lax 교수를 만날 기회가 종종 있었는데 그렇게 위대한 수학자인 그가 훌륭한 인품까지 겸비했다는 사실에 항상 감명을 받곤 했다. 현재 92세의 고령인 Lax 교수가 지금까지 남긴 수학적 성과들은 수학계 뿐만 아니라 물리와 공학의 여러 분야에서 아직도 매우 중요한 역할을 하고 있다. 그 중 하나를 예로 들자면 Lax의 등가정리(Equivalence Theorem)가 있다. 이 등가정리는 1956년, Communications on Pure and Applied Mathematics 라는 수학 저널에 처음 발표되었으니 매우 오래된 정리라고 할 수 있겠다. 이 논문에서 증명된 Lax의 등가정리를 증명하는 것은 그리 어렵지 않다. 그 후로 좀더 정밀하고 어려운 수학을 가지고 이 정리를 해석하고 증명하는 논문들이 많이 있었지만 1956년에 처음 나온 이 정리는 이해하기 쉽고, 쓰인 수학 역시 그렇게 어렵지 않다. 정리의 내용 역시 간단하고 이해하기 쉬운데 짧게 요즘 말로 해석하여 말하자면 다음과 같다.

주어진 미분 방정식을 수치적으로 풀어 해를 구한다고 할 때, 그 수치적 방법이 구간을 세밀하게 나누고 또 나누어 많은 양의 데이터가 쓰이면 더 정확해지는 방법이라고 할 때, 그 방법에 의해 얻어진 수치적 해의 안정성(stability)과 해의 수렴성(convergence)은 등가이다.

다시 말해서, 더 많은 양의 데이터를 쓰면 쓸수록 안정적인 수치적 방법은 해가 진리값(Truth)에 가까워지는 것을 보장해 준다는 말이다.

이 정리의 내용은 간단해 보이지만 매우 의미심장하다. University of Pennsylvania 에서 18,000 개의 진공관으로 만든 ENIAC이 등장한 것이 1946 년이고(1955년까지 사용되었다) von Neumann이 제안한 EDVAC이 나온 것이 1952년인 것을 생각하면 1956년에 발표된 이 등가정리는 향후 컴퓨터를 이용해서 얻은 수치적 해를 해석하는데 매우 중요하고 시기적절한 수학적 토대가 되었다. 이 정리를 바탕으로 공학자들은 마음놓고(?) 여러 물리적이고 공학적인 문제들의 해를 구하고 그 해가 참값에 가깝다고 결론을 내려왔다. Lax의 정리에 충실하여, 가능하면 더 많은 메모리와 CPU 를 사용하면서 매우 작은 스케일로 문제를 세분화하여 많은 양의 데이터를 이용해 진리값에 근접한 해를 구하려고 노력해 왔다.

그러나 그렇게 구한 해는 과연 얼마나 진리값에 가까울까. 만약 Lax의 정리가 맞다면 다소 걱정없이 수많은 데이터를 이용해 구한 해를 참값과 매우 가깝다고 말할 수 있다. 실제로 Lax의 정리는 다양한 문제들에 적용되어 수치적인 해의 정당성을 보장해 주었다. 아닌 게 아니라 컴퓨터의 성능은 Moore의 법칙에 의해서 기하급수적으로 진보하였고 방대한 양의 메모리를 사용하거나 병렬처리를 하는 것이 어렵지 않게 되었으며, 데이터의 양이 가히 천문학적으로 수집되는 현실에서는 많은 양의 데이터를 가지면 숨어있는 진리를 더 잘 발견할 수 있을 것이라는 기대는 당연해 보인다.

그런데 모든 수학의 정리들이 그렇듯이 Lax의 정리도 몇 가지의 가정을 전제로 하고 있다. 다시 말해서 만약 이 가정들이 맞지 않는 상황이면 Lax의 정리도 성립하지 않는다. 예를 들어서, 초기 조건이 존재하는 함수공간과 해가 존재하는 함수공간이 같은 공간이어야 한다거나 혹은 연산이 이루어질 때 무한대의 정확성(precision)을 갖는 식의 가정을 하고 있다. 두번째의 경우는 우리가 흔히 말하는 잡음(noise) 혹은 불확실성(uncertainty)에 관

한 가정이다. 수치적 해가 참된 해로 간다고 하면 그 해들이 존재하는 다양체(manifold) 공간에서 에너지가 가장 적게 소모되는 측지선(geodesic)을 따라 존재해야 하는데 불확실성이 존재하는 경우 이 수치적 해가 따라가는 측지선이 반드시 참된 해로 인도하는 측지선이 안 될 수도 있기 때문이다. 이 불확실성은 방정식 속에서 새로운 측지선을 만들어 낼 수도 있고 따라서 안정성의 문제는, 수렴성과의 등가 이전에, 구해진 해가 참된 해가 아닐 수 있음도 시사한다. 간단하게 말하자면 데이터의 크기가 크다는 사실이 그 데이터로부터 도출되는 지식이 반드시 참된 지식임을 보장할 수는 없다는 말이다. 정작 빅데이터를 얻고도 전혀 엉뚱한 결론에 도달할 수 있다는 뜻이다.

이런 이유로 IBM 은 빅데이터를 정의할 때 기존의 3V (Volume, Velocity, Variety)에 하나를 더 추가하게 되는데 그것은 Veracity 즉, 정확도에 관한 개념이었다. 이런 관점에서 보면 “The bigger the better” (데이터의 규모가 크면 클수록 좋다)라는 상식적 기대는 매우 위험한 구호일 수 있다. Lax 정리에서 보듯이 그 진리값을 발견해 내는 수학이 정밀하지 않으면 단순히 데이터의 크기가 크다고 해서 도출된 결론이 반드시 좋을 수만은 없다. 덩치가 큰 데이터가 그럴 듯하지만 정작 틀린 해석을 내놓거나 다른 목적으로 악용되는 경우 그리고 그것이 중요한 개인적, 사회적 결정과 사회적 안정성에 관한 문제와 연관되어 있다면 더욱 심각한 문제가 될 수 있다.

의심의 여지 없이 빅데이터가 화두인 세상이다. 큰 데이터이든 작은 데이터이든 그 안에는 자연의 이치와 사람들이 만들어내는 복잡 다단한 원리들이 숨어 있다. 그리고 데이터의 크기가 커지면 그 이치와 원리를 발견할 확률도 높아진다고 할 수 있다. 실제로 많은 영역에서 빅데이터의 혁신성과 가능성이 증명되고 있다. 그러나, 빅데이터는 로크의 경험주의식 ‘민방’에 들어와 ‘문자’와 ‘빛’으로 자동 변신하는 슈퍼지식 둔갑 로봇이 아니다. 그것이 적절한 지식으로 변화되어 사람을 살리고 사회 발전에 공헌하기 위해서는 그 변화를 이끌어내는 수학에 대한 꾸준한 관심과 연구가 병행되어야 함이 마땅하겠다.

특집

임베디드 시스템을 위한 빅데이터 분석 - 이젠 거꾸로! 기계가 하는 말에 귀 기울여야 한다.

전자공학과 이 정 원

프로그래밍 언어는 영어로 만들어진 인간의 말을 '0101'만 이해하는 기계에게 순식간에 수십억 개의 명령을 수행하게 만들었던 도구였다. 그러나 이제 거꾸로 프로그래밍 해야 하는 시대가 되었다. '거꾸로'라는 말의 의미는 기계가 하는 말을 인간이 이해할 때가 되었다는 뜻이다. 소프트웨어의 복잡도가 엄청나게 커지고 인간은 끝없이 기계가 이를 이해하고 빠른 시간 안에 수행해 내게 만들고 있다. 그러나 실제 기계가 이러한 명령을 수행해 내면서 만들어 내는 수많은 얘기들(예를 들자면 시스템 로그)에 이제는 귀 기울여야 할 때다. 필자는 이러한 거꾸로 프로그래밍의 기술을 '임베디드 시스템을 위한, 빅데이터 분석을 기반으로 하는, 프로그래밍!'이라고 본다.

인공지능 프로그램인 '알파고' 이후로 공학 분야의 딥러닝, 머신러닝, 빅데이터 분석 등의 기술도입은 실로 놀라울 정도로 확대되어 왔다. 필자도 2003년 데이터 마이닝을 주 기술로 박사학위를 받은 이후로 전자공학대에 재직하면서 임베디드 소프트웨어와 '인텔리전스'를 어떻게든 접목해 보려는 시도를 계속해 오긴 했지만, 2013년도에 아무런 전처리과정 없이 고양이 사진을 학습(learning)하면 동물 중에 고양이를 구별해 내는 딥러닝의 능력에 매료되어 바로 스테디를 시작하였다. 필자가 매력을 느낀 부분은 바로 '전처리(pre-processing)' 과정이 필요 없다는 점이었다. 데이터 마이닝¹⁾ 기술은 데이터 선정(Selection) → 전처리 → 변형(Transformation) → 마이닝(머신러닝) 알고리즘 적용 → 시각화 및 해석(Visualization&Interpretation) 단계를 거치는 아주 지루한 작업이다. 그런데, 딥러닝 알고리즘은 전처리 과정이 없다?! 스테디 결과, 딥러닝은 이미지를 학습하기에는 아주 잘 들어맞는다. 그러나 자율주행/전장 자동차나 산업용 로봇 등의 대표적인

임베디드 시스템을 학습하기에는 전처리 과정이 없어질 수는 없다.

마이닝 혹은 머신러닝 연구를 하면서 느낀 점은 전처리 과정의 복잡도와 정교함이 소프트웨어 엔지니어의 끈기를 시험한다는 점이었다. 본격적인 알고리즘의 적용은 오히려 재미있고 신난다. 머신러닝 알고리즘의 파라미터들을 이렇게도 저렇게도 설정하면서 기법을 정제해 나가는 작업은 어느 정도 눈에 보이는 작업이므로 시간 가는 줄 모르고 연구하게 된다. 그러나 전처리 과정은 기존에 엉망진창으로 noise와 outlier들이 잔뜩 숨어 있는 방대한 양의 저질 데이터를 어떻게 정제해야 의미 있는 예측(prediction)을 할 수 있을까? 고민하게 되며 정답을 알지 못하는 끝없는, 아주 지루한 작업이다. 예를 들어 아래의 산업용 로봇의 로그의 일부를 보자. 시스템의 상태를 끊임 없이 로그로 남기고 있다.

```
#STATE: CCtrlStateParallel ->CCtrlStateParallel
CCtrlStateParallel:: OnStateEntered: 1603: GPIO
SetOutput: loc(0), type(1), signal(0), index(2), value
(0.000000) #STATE: CCtrlStateParallel ->
CCtrlStateParallel CCtrlStateParallel::OnStateEntered:1603: Progress : 47/255
```

그러나 영역 지식(domain knowledge)을 모르고서는 각 로그의 용어(terminology)들과 연관관계(relationship)를 이해 할 수 없다. 한 마디로 이 기계가 얘기하고 있는 말을 전혀 이해 할 수 없다. 필자가 경험해 본 로봇은 한 달에 수 TB의 로그를 생성하고 자율주행 자동차는 하루에 2TB까지도 데이터를 생성해 낸다고 한다. 이러한 빅데이터 로그를 생성하는 문법을 파악하는 것이 임베디드 시스템을 위한 빅데이터 분석의 시작이다.

1) 데이터 마이닝(data mining): 대규모로 저장된 데이터 안에서 체계적이고 자동적으로 통계적 규칙이나 패턴을 찾아내는 것이다. 다른 말로는 KDD(데이터베이스 속의 지식 발견, knowledge-discovery in databases)라고도 일컫는다. 데이터 마이닝의 핵심 단계에서 머신러닝 알고리즘의 선택은 마이닝의 정확도를 결정한다.

그런데, 이걸 우리가 새로운 프로그래밍 언어를 배울 때와 똑 같지 않은가? 용어를 이해하고 문법(syntax)을 배운다는 것이... 웹 데이터나 이미 구축된 데이터베이스로부터의 빅데이터 분석은 오히려 쉽다. 그러나 임베디드 시스템에서의 빅데이터 분석은 전처리가 매우 어려운 작업이다.

필자는 최근 로봇의 예지 정비(Predictive Maintenance)를 위해 빅데이터 분석기법을 적용하는 연구를 진행하고 있다. 이때, 단순히 로봇의 토크값, 전류/전압값, 속도 및 각도, 위치 등의 숫자를 수치로 분석하기 보다는 이 정보들이 기계가 무슨 얘기를 하고 싶어 하는 건지로 바꿔 해석해 보는 작업을 하고 있다. 이러한 모든 파라미터들의 값을 조합하고 인간이 내렸던 명령어와 연관지어 생각하면, ‘제(기계)가 B 라는 명령을 수행하려다 보니, 3 축의 힘(토크값)이 자꾸 시간이 갈수록 더 들어가네요. 그러니 내리신 명령에 맞는 속도를 매달 1.2%씩 감소되어 맞출 수가 없어요’라고 로봇이 얘기하고 ‘음.....그렇다면 B 보다

전에 내린 A 명령어에서 너무나 급격한 변화가 원인이군... A→A’→B 로 서비스를 변경하는 것이 로봇에 무리가 없겠어’ 라는 시나리오가 가능하다.

필자는 앞으로 5년 이내에 기계가 하는 말을 알아듣기 위한 임베디드 시스템을 위한 빅데이터 플랫폼 개발이 필수적일 것이며, 각 도메인마다 자신들이 필요한 해석을 하기 위해 센싱값, 로그값에서부터 명령어, 환경정보, 서비스 레벨까지 잇는 새로운 패러다임의 거꾸로 프로그래밍 언어와 API가 개발되어야 함을 확신한다.

특집

생물정보학 및 유전체학 최신 동향

생명과학과 박대찬

생물정보학을 이용한 유전체학 연구가 각광을 받고 있는 이유는 인공지능 빅데이터 시대의 “정밀의료” 열풍과 맞닿아 있다. 우선, “정밀의료”에 대한 정의를 살펴보자. 정밀한 의료는 단어 자체로 세련되고 좋아 보이지만, 어떤 의료를 정밀하다고 정의할 수 있는지는 모호하다. 오바마 정부의 Precision Medicine Initiative(PMI)에 따르면, genome sequence, microbiome composition, health history, lifestyle, diet를 정밀의료 구현을 위한 정보로 규정하고 있다. 즉, 분자 수준에서 측정하는 오믹스 데이터(Genomics, Proteomics, Metabolomics, Microbiome 등) 분석에 대한 필요성은 학문 분야를 넘어서 우리 실생활에 스며들고 있다.

오믹스 기술 중 시장 및 연구 규모가 가장 큰 분야는 단연 Genomics(유전체) 분야이다. 이 유전체 데이터를 생산하는 기술로 차세대염기서열분석(NGS)이 널리 이용되고 있다. NGS는 뒤섞인 DNA 조각을 동시에 서열을 밝히는 기술로서, 현재 많이 사용되는 NovaSeq이라는 기계는 약 2000 giga base의 데이터를 1주일에 생산할 수 있다. Human genome sequence가 3 giga base이며 1990년 - 2000년 초 human genome project가 10년이 걸린 것을 고려했을 때, 현재 생산되는 데이터의 크기와 생산 속도는 천문학적으로 증가하였다. 이제는 어떻게 데이터 빠르게 많이 생산할 것인가보다 어떻게 데이터를 분석하고 보관할 것인가에 대한 생명정보학적 질문이 더 큰 고민거리가 되었다.

▶ 5페이지에 계속

이러한 오믹스 데이터 연구는 컨소시엄을 통해 대규모화되는 추세이다. 정밀의료를 구현하기 위한 빅데이터를 생산하기 위해서는 여러 연구기관 또는 국가들이 협력하여 데이터를 생산할 필요가 있기 때문이다. 유럽, 미국에서 진행되는 대규모 암연구 컨소시엄인 The Pancancer Analysis of Whole Genomes(PCAWG), The International Cancer Genome Consortium(ICGC), The Cancer Genome Atlas(TCGA)에서 생산된 데이터가 통합 분석되어 2018년 공개 되었으며, 이는 암환자 4,645명의 whole genome sequencing(WGS), 19,184명의 whole exome sequencing(WES)에서 찾은 84,729,690개의 돌연변이를 보고 하였다. 이 중, 2005년에 시작한 미국의 TCGA 프로젝트만 보아도 지금까지 1만 1천여 명의 암 환자에서 2.5 Petabytes 이상의 데이터를 생산하였으며, 이 데이터를 저장할 수 있는 국내 연구기관은 1~2곳에 불과한 실정이다.

국내 대부분의 유전체 데이터 생산은 business-to-business sales(B2B) 방식이다. B2B는 병원 또는 학교에서 생체 시료를 모아 유전체 생산 회사에 데이터 생산을 의뢰하는 방식을 의미한다.

위에 소개한 바처럼, 미국의 경우도 정부주도 연구과제를 통한 B2B의 형태로 질병 유전체 데이터가 대규모로 축적되기 시작하였다. 하지만, 앞으로 유전체 빅데이터는 Direct To Customer(DTC: 기업에서 개인에게 직접 제공되는 방식)에 의해 수집될 것으로 예측된다. 미국의 사기업인 23andMe가 대표적인 DTC 유전자 분석 기업이며, 이미 100만 명 이상의 유전자 정보를 수집하였다.

유전체 데이터의 경우, 전 세계적으로 작년까지 이미 200만 명 이상의 WGS / WES 데이터가 생산 되었을 것으로 추정되며, 2025년까지는 약 10억 명의 유전체 데이터가 축적될 것으로 전망한다. 따라서, 짧은 미래에 민간 기업에서 대량의 유전체 데이터를 경쟁적으로 확보하게 될 것이며, 큰 데이터를 분석할 수 있는 인프라를 구축하는 것이 주요해될 것이다. 또한, 유전체 변이 발굴을 위한 생명정보학 기술을 넘어서, 임상 및 생체 정보(질병 치료 기록, 식습관, 혈당, 심박수)를 포함시킨 기계학습법이 생명정보학 분야를 주도할 것으로 보인다.



디지털역사학의 시작과 도전

사학과 이상국

연구실 창 밖으로 그려지는 사계를 사랑한다. 봄의 소생과 여름의 열정, 가을의 낭만과 겨울의 암울. 자연의 넉넉한 품 속에서 겨울의 암울마저도 사랑스럽다.

상아탑은 ‘변화’와 ‘위기’로 물들어 있다. 상아탑의 ‘가치’는 ‘사치’로 치부된 지 오래다. 철새 없이 몰아치는 변화의 물결은 그 파고를 더해가며 압박해온다. 상아탑의 품은 낭만마저도 넉넉히 품을 수 없다.

사회는 ‘혁명’으로 팽팽한 긴장감을 더해간다. 차수를 더해가는 산업혁명은 더욱 더 알 수 없는 개념과 신조어들을 만들어내며 그 만큼의 불안과 위기의 강도를 높여간

다. 사회는 더 이상 다양성의 용광로가 아니다. 벌어진 판에 안착한 자만을 위한 사회이다.

역사적으로 보면, 이 모든 현상은 그리 특별한 일만은 아니다. 인류가 위기 아닌 때가 얼마나 있었으며, 변화를 강요 받지 않았던 때가 언제 있거나 했던가. 상아탑에 닥친 위기와 변화는 시나브로 스며들어 이제 만연하다 못해 식상할 정도가 아닌던가. 난마처럼 영겨있는 문제도 의외로 단순하고 본질적인 곳에서 실마리를 찾아 해결해오곤 했다.

▶▶▶ 6페이지에 계속

사회가 아무리 변해도 진리의 탐구와 미래 청사진에 대한 소임을 맡을 곳은 상아탑이 아니던가. 우리 대학만은 그래도 되지 않겠는가. 아니 우리 대학만이라도 그래보지 않겠는가. 현실에 대한 냉철한 인식을 바탕으로 상아탑의 본질에 충실하는 것 말이다.

그래서 ‘디지털역사학’을 시작한다. 역사학을 좋아하는 학생에게 졸업 후 직장을 얻는데 용이하게 하기 위해서라기 보다는 졸업 후 살아갈 인생을 스스로 설계할 수 있게 하기 위해서 말이다. 역사학의 본질에 한 걸음 한 걸음 나아가면 갈수록 변화하는 사회에 주체적으로 대응하는 능력을 갖게하기 위해서 말이다. 앞으로 더해질 더 많은 차수의 산업혁명에도 두려움보다는 기회의 장으로 활용할 수 있는 역량을 키우게하기 위해서 말이다. 역사학의 전문 연구자들에는 양적연구(quantitative analysis) 방법론으로 안내해 역사연구 영역의 확장을 유도하기 위해서 말이다. 다양한 학문분야의 방법론을 방대한 ‘역사학 빅데이터’에 적용하는데 필요한 기준을 제시하고, 사료의 역사성에 바탕을 둔 해석의 방향성을 정립하는 것 말이다. 사회에는 융복합이라는 시대적 화두에 역사학이 중심이 되어 응답하고 그 사회적 책임을 다하는 것 말이다.

역사학의 본질은 변화된 사회와 연구 환경에 적응하지 못하여 스스로 고립을 자초하는 것이 아니라, 끊임없이 질문을 던지고 최적화된 방법을 찾아 알맞은 해답을 탐구해가는 것이리라. 사실[진리]를 탐구하는데, 사료의 자구 하나하나 그리고 그 행간의 의미까지도 세세하게 해석하는 것이 옳다면 그래야 하며, 방대한 ‘역사학 빅데이터’를 적절히 활용하기 위해 텍스트마이닝 기법으로 필요한 정보를 추출하고 데이터베이스화해서 통계분석이나 시각화해야 할 필요가 있다면 이도 또한 주저하지 말아야 한다. 기존의 실증적 역사학 연구방법론이 여전히 유효하고 필요하다면 계승해야 하며,ダイナ믹하고 장기적인 역사적 사실과 추세를 분석하는데 데이터마이닝 기법과 사회연결망 분석방법, 머신러닝과 딥러닝 기법, 그리고 위상수학 분석방법이 요청된다면 저어함이 없어야 한다. 기본에 충실하다는 것은 이전의 익숙함에 머무르는 것이 아니라 변

화된 환경에 적응하여 끊임없이 확장하는 기초체력을 축적한다는 것을 의미하는 것이 아니겠는가.

그래서 역사학에 ‘디지털기술’을 더한다. 사회과학에서 발전시켜온 통계적인 방법론을 적극적으로 수용하고 적용하며, 역사학의 기본인 사료를 데이터의 측면에서 분석하는 데이터사이언스와 수학을 포함한 공학과 자연과학과의 만남에도 적극적이하고자 한다. 이렇게 역사학의 본질에 대한 천착에서부터 시작한 디지털역사학은 분석과 해석 능력을 바탕으로 교육의 영역으로 확대된다. 사회에 나아가기 전 일생의 틀을 마련하여 과거와 미래에 대한 책임을 다하는, 기본기가 튼튼한 미래형 인재를 육성하는 상아탑 본연의 가치를 회복한다.

“대학의 본질은 단순히 현재의 문제에 책임을 지는 것이 아니라 과거와 미래에 대해 책임을 지는 것입니다. 대학은 다음 학기에 나올 결과나 학생들이 졸업한 뒤 갖게 될 직업만을 다루는 곳이 아닙니다. 일생의 틀을 마련하고 수천 년의 유산을 후세에 전하는 동시에 미래를 구체화하는 배움터입니다.”(하버드대학교 총장 드류 길핀 파우스트(Drew Gilpin Faust)의 2007년 취임사에서, 아주대 사학과 홈페이지 학과장 인사말 재인용)

이제 역사학은 모든 학문분야를 아우르는 종합학문으로서 학문적 위상을 회복하게 된다. 진리를 탐구하는 상아탑의 가치는 다시 넉넉한 품으로 다양성의 용광로에 풀무질하게 된다. 사계절을 품은 소소한 삶은 상아탑에 대한 자부심으로 인간과 사회에 대한 무한한 책임과 신뢰로 이어진다. 디지털역사학의 도전이 시작되었다.

특집

의학과 데이터과학의 만남

의학과 의료정보학과 윤덕용

나는 의사(정확히는 의사 면허증 소지자)이자 데이터 과학자이다. 아직까지는 흔하게 볼 수 있는 조합은 아닌 듯 하다. 나의 이런 복합적인 경력 덕분에 의료계에도, 컴퓨터 과학계에도 노출될 기회가 종종 있는데, 이 두 사회를 겪어보니 왜 이런 조합이 쉽지 않은지 이해하게 된다. 의사들의 사회는 모든 과학분야 중에 가장 경직되고 보수적이라는 데에 대부분 동의할 것이다. 새로운, 창조적인 시도가 검증되지 않았다면 오히려 환자들에게 해를 끼칠 수 있다는 점을 생각하면 충분히 이해할 수 있다. 그래서 의학도들에게는 수업시간에도 기존의 지식을 습득하는데 익숙하고, 알려진 대로 그대로 말하고 행동하는 것이 당연하다. 하지만 컴퓨터 과학 분야는 모든 과학 분야 중에 가장 자유로운 분위기가 아닐까 생각된다. 사실 그들은 창조적이지 않으면 생존하기가 힘들다. 이는 수업시간에서도 나타난다. 한번 정보통신대에서 특강을 할 기회가 있었는데, 그 수업 분위기는 너무나 충격적이었다. 의대 수업에서는 수업시간 전 모두가 출석하여 바르게 앉아 있는 모습이 기대되었는데, 정보통신대의 수업에서는 모두가 자유로워 보였고 교수와의 의사소통이 너무 편하게 느껴졌다. 이런 느낌을 갖고 수업을 마쳤는데, 초청해 주신 분의 마지막 인사가 '오늘 의대 교수님이 오셔서 다들 평소보다 진지하게 수업에 참여한 것 같다'여서 한번 더 충격을 받았던 기억이 있다. 학회에서의 분위기도 많은 차이를 보인다. 의학 분야 학회에서는 모두가 반듯하게 정장을 입고 오는데, 컴퓨터 과학 분야 학회에서는 자유로운 복장이 대부분이다. 양쪽 학회를 오가다가, 한번씩 드레스코드를 잘못 맞춘 경우에는 상당히 이질적인 시선을 느끼게 된다.

그런데 이렇게 이질적인 의학 분야와 데이터 과학 분야가 서로 만나기 위한 몸부림을 치고 있다. 지난 11월 2일부터 4일까지 Korea Clinical Datathon 2018 행사가 있어 참석하였다. Datathon이란 data와 marathon의 합성어로,

하나의 임상적 문제를 풀기 위해 2박3일간 쉬는 시간 없이(심지어 자는 시간도 따로 없다) 팀원들이 협력하여 데이터 분석을 수행한다. 이번 행사의 목적은 실질적으로 의미 있는 분석 결과를 도출하는 것이 아니라, 의사와 데이터 과학자를 서로 "만나게" 하는 데에 있다. 한 팀은 보통 의사 2-3명, 데이터 과학자 2-3명, 그리고 일반인 참가자(주로 학부생) 2-3명으로 구성되었다. 의사들은 현장에서 필요한 임상적 문제를 제시하고, 그 문제가 무엇인지, 왜 중요한지를 데이터 과학자와 일반인 참가자들에게 설명한다. 데이터 과학자는 어떠한 방법으로 데이터를 처리하고 분석하는 것이 해당 문제에 가장 적합한지 제시한다. 처음에는 서로를 이해하는 것이 완전하지 않기 때문에, 많은 시행과 착오를 경험하게 되고, 그 과정을 통해 서로를 이해하게 된다. 이번 행사에서는 10개 팀이 구성되었고, 모두가 2박 3일 동안 datathon을 하였다. 이러한 행사는 현재 전세계에서 진행되고 있다.

우리 아주대학교에서도 이러한 움직임을 확인할 수 있다. 아주대학교 의과대학에는 내가 속한 의료정보학과가 있으며, 교수와 학생 모두 의사와 데이터 과학자의 조합으로 구성되어 있다. 나와 박래용 교수님은 의사 출신이고, 박범희 교수님은 통계학을 전공하였다. 학생들의 출신 학과를 보아도 의학, 약학, 생물학, 컴퓨터 과학, 전산, 산업공학, 통계학 등 의료 데이터를 이해하고 분석하는데 필요한 다양한 학문 분야가 모여있다. 이렇게 모인 인원이 어 느덧 50여명에 달하고 있다. 그리고 우리는 2박 3일이 아닌 10여 년째 datathon을 하고 있다.

왜 이렇게 이질적인 두 분야가 만나기 위해 힘든 노력을 하는 것일까? 가장 쉽게 생각할 수 있는 이유 중 하나는 병원에 데이터가 많기 때문일 것이다. 사실 의학은 그 시작부터가 데이터에 기반을 둔다. 의사는 환자의 증상을 묻고, 직접 관찰한 내용을 기록해왔다. 그리고 그 기록을

바탕으로 환자를 치료한다. 그리고 컴퓨터가 도입되면서 많은 데이터들이 컴퓨터 내에 저장되기 시작하였고, 어느 정도 시간이 지난 지금, 데이터에 목이 마른 여러 데이터 과학자들의 관심을 끌기에 충분히 많은 데이터가 진산화되어 쌓이게 되었다. 또 다른 이유 중 하나는 의료 현장에서 다루는 생명 현상의 복잡성이 매우 도전적이기 때문일 것이다. 데이터가 아무리 많더라도 풀고자 하는 문제가 매우 단순하다면 데이터 분석의 가치가 높지 않을 것이다. 하지만 병원에서 발생한 데이터에는, 수많은 경우의 수가 존재하며, 또한 아직 우리 인류가 풀지 못한 수많은 난제들이 포함되어 있다. 이러한 복잡한 의료 데이터의 과학적 분석을 통해 새로운 의학 지식이 발굴되면 인류의 건강상태가 증진될 것이고 또한 의료비용을 감소시키는 등 많은 실질적 이득을 줄 것으로 기대된다. 따라서 국가적으로도 사회적으로도 의료와 데이터 과학의 조우가 환영 받고 있다.

하지만 이 둘은 서로 다른 세계에서 만나기 위해 모이고 있기 때문에, 그 간극을 줄이기 위해 풀어야 할 오해가 산적해 있다. 현재 가장 문제가 되고 있다고 생각하는 부분은, 서로에 대한 불신이 아닌, 오히려 서로에 대한 과대한 기대라고 생각한다. 의사들은 인공지능이 모든 문제를 해결해줄 것이고, 더 나아가 의사를 완전히 대체하게 될 것이라는 생각을 갖고 있다. 이는 비단 의사들만의 생각이 아니라, 현재 공고되어 나오는 국가 프로젝트 목록을 보아도 사회적으로 얼마나 과도한 기대를 걸고 있는지 쉽게 알 수 있다. 반면 데이터 과학자들은 병원에 쌓여있는 데이터들이 분석에 바로 쓸 수 있을 만큼 잘 구조화되어 있고 오류가 많지 않을 것이라고 기대하고 있다.

의료계에는 현재 인공지능 기술의 수준을 이해하고, 그 수준에 맞는 문제에 기술을 적용해야 한다는 인식이 필요해 보인다. 당장 의사를 대체할 수 있는 수준의 기술은 아니더라도, 의사의 진료나 치료 행위를 작은 범위라도 실질적으로 도와줄 수 있는 기능이 현 단계에서는 더욱 큰 영향을 줄 수 있을 것이다. 비록 다소 단순한 작업일지라도 컴퓨터는 지치지 않고 의료진을 지원해줄 것이고, 나아가

기술의 발전에 따라 그 수준이 점점 증가하게 될 것이기 때문이다. 다른 한편으로는 수집하고 있는 데이터에 대한 보다 진지한 고민이 필요하다. 현재 운영되고 있는 대부분의 병원정보시스템, 전자의무기록 시스템 등은 데이터 분석에 대한 준비가 부족하다. 실제 진료 현장에서 데이터를 잘 보여주지만 하는 데에 목적을 두다 보니, 내부적으로는 데이터가 표준화되어 있지 않거나 변경 이력들이 제대로 남아있지 않은 경우들이 자주 있다. 또한, 현재 모이고 있는 데이터의 규모가 양질의 분석 결과를 도출하기에 충분하지 않다는 점도 많은 의료진들은 인지하고 있을 것이다. 지금 당장 진료에 중요하지 않은 변수들은 처음부터 수집되지 않았으며, 진료 및 치료 시 수집된 많은 데이터들이 최종적으로 병원 데이터베이스에 저장되어 보관되지 않은 경우도 많다. 어떠한 데이터들을 추가적으로 수집할 수 있을지, 그리고 지속적으로 수집하기 위해 어떻게 수집 과정을 자동화할 수 있을지를 고민해야 할 것이다.

사실 의학이라는 분야는 고대부터 데이터에 기반해왔으나, 과학이 발전되는 과정에서 세분화되면서 이 둘은 각자의 길을 가게 되었고, 눈부신 발전을 이루었다. 하지만 이제 다시 학문간의 경계가 희미해지고 도처에서 융합학문의 필요성이 대두되고 있는 시점에서 서로의 간격을 다시 줄이고자 하는 노력이 필요하게 되었다. 이러한 노력이 전세계 각지에서 치열하게 이루어지고 있는 만큼, 의학과 데이터 과학의 만남은 인류에 큰 기여를 할 것으로 기대된다.

소통과
담론

교원 임면권 관련 아주대학교 정관 변경에 대한 우리의 우려

지난 8월 23일 학교법인 대우학원 이사회에서는, 종래 총장에 위임됐던 교원의 임면권을 이사장이 행사하는 것을 골자로 한, 정관 변경을 의결하였다.

교수회에서는 올해 두 차례 성명을 통해, 이사장에게 권한 집중의 문제점을 지적하였다. 이런 상황에서 교원의 임면권마저 이사장에게 넘기는 것은 총장의 권한을 크게 제한할 것이며, 대학의 자율적 발전에 걸림돌이 될 우려가 클 뿐 아니라, 학내 자율성을 기반으로 발전해 온 우리 학교의 명예로운 전통을 훼손할 가능성이 있다. 이런 점에서 교수회는 이사회에 일방적인 정관 변경을 반대하는 입장을 표명하였다. 그러나 이사회는, 교수들의 충정을 무시하고 교수회의 의견에 아무런 설명도 없이, 독단적으로 정관 변경을 의결하였다.

이에 우리는, 학교 발전을 위한 교수회의 정당한 의견이 아무런 해명 없이 재단 이사회에서 무시당하는 아주대학교에서의 비민주적 운영 행태와 이 과정에서 드러난 총장의 역할에 심각한 우려를 표하지 않을 수 없다. 추후 교수회에서는 이를 바로 잡기 위한 노력을 경주할 것이다. 또 이러한 의사결정 과정에 참여한 총장은 학내 구성원들에게 양해를 구하고, 변경된 정관에서 대학의 자율성을 최대한 보장하기 위한 제도적 장치를 마련하여야 한다.

2018년 9월 19일

아주대학교 교수회

우리병원 남자들 관찰기

의학과 감염내과학교실 최영화

우리 병원에 남자들이 좀 있는데, 모 여선생이 그들의 이름을 직접 쓰지는 않고(왜 그랬을까?) 번호를 붙이기 시작한 것 같습니다. 이 남자들을 좋아했거나 사랑했는지 아니면 그저 관찰만 했는지는 알 수 없고 어쨌든 누구에게 보내는 편지인 듯, 소개서인 듯 몇 자씩 적어놓았는데 어느 날 우연히 화장실에 놓고 간 것을 내가 발견했지요. 긴 시간 머물기 좋아하는 후천성 배변장애인인 내가 스스로 나가지 않는 한, 주인이 찾으러 와도 소용없다는 것을 알아채고, 어쨌든 안 읽었다고 말하기로 마음먹고 읽기 시작했습니다. 물론 그이의 글은 제법이었고 더불어 나 또한 남자에 관심이 많았기에 이 번호 붙은 남자들이 누구인지 궁금하고 또 나중에 심심할 때 찾아볼 요량으로 살짝 찍어두었는데(요새는 늘 사진기가 손에 있기로) 이 자리에서 그냥 써먹기로 했다.

001. 이 남자는 손은 외과인데 다른 모든 것은 내과입니다. 내리 긴 시간 회진을 돌고 꼼꼼히 진단하고 늘 궁리하지요. 무엇 때문인가? 요즘엔 집에 가긴 가는데 그래도 상당 시간 병원에 있습니다. 저는 제게 무슨 문제가 생기면 이 의사한테 입원하고 싶습니다. 믿고 하라는 대로 하고 이 사람이 내가 죽을 거라고 하면 그럴 수밖에 없다고 생각하고 이제 그만 가서 쉬라고 얘기해주고 싶습니다. 중환자실 불을 꺼준 것에 대해서도 감사하다고 말하겠지요. 수술을 하지만 수술 이후의 문제를 깊이 숙고하는 의사입니다. 수술 이후의 문제 이것이 내과이고, 수술만으로 해결할 수 없는 인체의 문제가 내과의를 늘 괴롭히는데 이 남자는 그걸 압니다. 그래서 제가 보기에 그는 손만 수술하고 나머지는 다 우리 편 내과라는 생각입니다.

002. 그는 사실 밖에 나가면 매우 비싼 의사입니다. 사실 그가 쳐다볼 곳은 다른 곳일 수 있었습니다. 그런데 남들이 다 보는 것을 피해서 엉덩이와 다리만 보게 되는 운명을 받아들이는 것 같습니다. 이쪽 살을 저쪽에 붙여주고

다 덮어주지요. 그는 얼굴로 기억하지 못하고 상처와 상처를 케맨 흉터로 환자를 기억하는 슬픈 운명을 팔자로 여기는 것 같습니다. 나는 그가 괴팍한 환자들을 피하지 않고 받아서 곱게 다듬고, 그들이 가진 이런저런 합병증을 모르는 것인지, 안 무서워하는 것인지 어쨌든 피하지 않는 것에 경이감을 느낍니다. 사실 그런 그를 보조하는 아랫사람들은 환자만큼 그도 괴팍하다고 합니다. 그의 괴팍함이 우리의 뒷배입니다.

003. 그가 원래는 그렇게 초췌하지 않았다는 것을 저는 알고 있습니다. 그는 깍은 듯한 남자입니다. 그런 사람이 삭아가는 것을 보면 환자 보는 일이 참 힘든건 힘든 일이란 생각을 합니다. 수술 이후의 문제를 다루는 것이 원래 그의 문제는 아니었을 겁니다. 그러나 수술 이후의 숨이 찬 문제가 그를 괴롭히는 주요 문제입니다. 왜 사람은 그렇게 단순한 기계가 아닌가요? 왜 수술은 늘 잘되고, 그 이후의 고통은 늘 내 몫인가? 이렇게 부르짖을 만한데 그 또한 숨이 찬 고통 앞에서 물러서지 못하는 사람입니다. 새 가슴이라 그렇지요. 그가 새가슴이라 누군가는 발 뺀고 잠을 잘 것입니다.

004. 그는 가르치는 것에 희열을 느끼는 사람입니다. 물론 그의 가르침은 너무 깁니다. 만연체라서 듣다가 졸거나, 주어가 원래 무엇이었는지를 모르게 되거나, 걸어다니는 교과서라 너무 고리타분하게 느껴지기도 합니다. 그의 긴 설명 중에 입에 침이 마르는 것을 보는 것은 흔한 일이지요. 그렇지만 그는 예비 의사들에게 무엇을 어떻게 가르쳐야 하는지 아는 사람입니다. 없는 방식을 스스로 만들고 만들어서 실천하는 가르치는 일의 강자지요. 또 그는 어떻게든 남의 환자를 좋게 만드는 데 혈안이 되어 있는 사람입니다. 밤늦게까지 남 좋은 일 하는 게 그렇게 그에게 중요한 일이냐고 물으면 그렇다고 할 것입니다. 젠장.

005. 그는 오전 8시 즈음에 아침 회진을 시작합니다. 회진은 고도의 집중을 요합니다. 한 눈 팔면 금방 딴 길로 새지요. 위아래 합심하여 집중하지 않으면 안 생길 일도 생깁니다. 매일같이 되풀이되는 이 일상, 남들 보기에 당연한 이 하찮아 보이는 병동 뱅뱅 돌기가 그가 하는 일의 중심입니다. 그러고는 외래를 보지요. 일주일에 네 번은 해야 합니다. 그리고 또 시술이 있습니다. 오전 오후 타과에서 부탁하는 잡스러운 일을 해결해줘야 합니다. 잡스럽게 한 줄을 쓰기 위해서 사소하게 앉아 파악하는 시간은 또 잡스럽게 흘러갑니다. 일상의 무거움, 임상의 무거움입니다. 그도 세월이 좋을 때는 이렇지 않았답니다. 평생 이렇 줄은 몰랐는지요. 오호 애재라. 그래도 그냥 버티는 것은 자부심 때문입니다. 저한테 오셔서 다행이에요. 적어도 성의 없는 허튼 진료는 안 받으실 겁니다. 그런 생각을 하지요. 물론 때로 저도 우울하긴 해요. 술 한잔에 내비치는 마음입니다.

006. 그는 자기가 교수가 된 것이 운이 좋아서라고 만방에 소문을 냈습니다. 앞으로 이십 몇 년간 정년을 보장한다는 문서를 받았다고 자랑도 했지요. 앞으로는 자기처럼 운 좋게 교수가 되기는 어려울 거라고 예언을 했습니다. 그는 자기 업적을 겸양으로 말하는 것이었고 험해지는 교수사회를 말하고자 하는 것이었는데 정말 돛자리를 깔아도 될 사람입니다.

007. 그가 이런 말을 했습니다. 몇 개의 창이 빛난다고 해서 그 창만으로 아주대 병원이라는 성을 지킬 수는 없습니다. 진료로 이루어진 성이고 이 성에는 내가 아는 주춧돌이 많습니다. 물론 내가 모르는 굳센 버팀목들도 여기 저기 있겠지요. 작거나 크거나 눈에 띄거나 그렇지 않거나 이 기반이 모여서 쌓아지는 미래요, 이루어온 과거입니다. 묵묵히 있는 이들의 자부심을 흔들거나 비웃거나 알잡아 보는 일이 없기를 바랍니다. 성을 쌓기는 힘들어도 무너뜨리기는 쉬우며 무너진 후에 다시 쌓기는 처음 같지 않을 테니까요.

나는 묵직한 뒤를 해결하고 나오면서 이 007이 누구일까 궁금했다. 001에서 006까지는 제법 나의 더듬이를 굴려서 짚어볼 수 있었는데 이 007은 참, 누더 나온 것처럼 찻찻한데 한 번 얼굴이라도 보았으면 싶습니다.



소식

교수회 소식

교수회 제91,92,93차 월례 대의원회 개최

지난 9월 10일 오후 6시 종합관 12층 코너스톤에서 제91차 월례 대의원회가 개최되었습니다. 보고사항으로 신임대의원 선임과 관련된 사항을 보고하였습니다. 논의사항으로는 교원 임용권 정관 개정에 따른 대응방안 논의, 교수회 부의장 재선임 및 차기 교수회 의장 및 감사 선출과 관련된 내용, 교수회 가을 산행 일정 등을 논의하였습니다.

또한 지난 10월 22일 오후 5시 종합관 12층 코너스톤에서 제92차 월례 대의원회가 개최되었습니다. 보고사항으로 차기 교수회 의장 및 감사 선거 일정을 논의하였으며, 광고산 가을산행 일정 관련 사항을 논의하였습니다.

그리고 지난 11월 12일 오후 6시 종합관 12층 코너스톤에서 제93차 월례대의원회가 개최되었습니다. 보고사항으로 차기 교수회 의장 및 감사 선거 진행 상황 및 교수회 홈페이지 제작 관련 사항을 보고하였습니다. 논의사항으로는 추후 진행 예정인 차기 교수회 의장 및 감사 선거 관련 내용을 논의하였습니다.

교수회에서는 매 학기 월례 대의원회를 개최하고 있습니다. 앞으로도 교수회의 활동에 많은 관심 부탁드립니다.



편집위원회: 김종식(편집책임),
김상배, 노명우, 윤호섭, 이재호