

한의학 연구자를 위한 시스템 생물학 학습 가이드

김창엽*

가천대학교 한의과대학 생리학교실

Guide to Learning Systems Biology for Korean Medicine Researchers

Chang-Eop Kim*

Department of Physiology, College of Korean Medicine, Gachon University

The emergence of systems biology in the 21st century is changing the paradigm of biomedical research. Whereas the reductionist approaches focus on components rather than time or contexts, systems biology focus more on interrelationships, dynamics, and contexts. The key ideas of the systems biology shares much with the philosophy of Korean Medicine(KM) and therefore, the paradigm shift is shedding light on understanding the mechanism of action of KM at system level. In this article, I provide a guide to learning systems biology for KM researchers using online learning resources. Thanks to the recent development of MOOC(massive open online courses) and other online learning platforms, learners can access to plenty of high-quality resources from top-tier universities in the world. I expect this guide help researchers to employ systems biology methods into their KM researches, and will lead to the development of future curricula for training "bi-lingual" experts, KM and computational approaches.

keywords : Systems biology, Korean Medicine, Online learning, MOOC

서론

21세기에 접어들어 다양한 대용량 기술(high-throughput technology)의 발전과 함께 생물학 분야가 본격적인 빅데이터 시대로 진입하였다. 여기에 시스템 과학(systems science)의 방법론들이 적용되면서 생물학은 기존의 환원주의적 접근을 넘어 전체로서 생명현상을 이해하고자 하는 시스템 생물학(systems biology)의 시대로 진화하고 있다¹⁾. 그리고 이러한 생명과학/의과학 분야의 패러다임 변화와 새롭게 개발된 방법론들은 전일론적 접근(holistic approach)을 지향해왔던 동양의학을 재조명하게 하는 계기가 되고 있다²⁾.

중의학계는 2000년 이후 시스템 생물학의 발전에 발맞춰 새로운 방법론들을 중의학 연구에 적극적으로 도입하여 왔으며, 변증(辨證), 방제(方劑) 등의 분야에서 과거 환원주의 일변도의 연구들과는 차별화되는 성과들을 빠른 속도로 내어놓고 있다³⁻⁵⁾. 미시적 수준의 생물학적, 정량적 데이터(quantitative biological data at micro level)에 시스템 수준의 분석과 모델링(analysis and modeling at system level)을 적용함으로써, 구체성을 바탕으로 거시적 이해를 추구해 나가고 있는 것이다. 전일론적 철학과 경험을 가진 동양의학에 환원주의 과학이라는 맞지 않는 옷을 입히려

애썼던 것이 20세기까지의 과학화, 현대화 연구였다면, 환원주의의 구체성과 시스템 과학의 전일성을 겸비함으로써 보다 몸에 맞는 과학의 옷으로 갈아 입을 수 있게 된 것이 21세기의 새로운 연구 동향이라고 표현할 수도 있을 것이다.

최근 국내 한의계에도 이러한 연구동향에 관심을 갖고 있는 연구자가 늘어나고 있다. 그러나 기존의 실험생물학적 연구방법론과 달리 수학/통계/컴퓨터 관련 지식과 기술이 우선적으로 필요한 시스템 생물학 연구의 특성상, 아직 연구를 진행할 수 있는 기술과 역량을 갖춘 한의학 연구자는 소수에 불과한 것이 현실이다. 뒤쳐진 한국 한의계의 관련 연구 역량을 강화하기 위하여 우선 한의학 연구자와 시스템 생물학 전문가 사이의 협력 연구가 활성화되어야 하겠으나, 장기적으로는 한의학을 중심으로 시스템 생물학의 다양한 방법론을 적용할 수 있는 융합형 연구인력 양성이 반드시 필요하다. 그리고 이를 위해서는 한의학 전공자를 대상으로 정량적, 계산적 방법론을 교육하기 위한 커리큘럼도 필요할 것이다.

본 논문에서는 저자의 경험을 바탕으로, 시스템 생물학의 연구방법론을 공부하고자 하는 한의학 전공자들에게 학습 가이드를 제공하고자 한다. 이를 통해 보다 많은 한의학 전공자들이 새로운 연구 흐름에 진지하게 참여할 수 있을 것이며, 나아가 양 분야의 언어에 모두 능한(bi-lingual) 한의학 연구자를 양성하기 위한 커리큘럼 개

* Corresponding author

Chang-Eop Kim, Department of Physiology, College of Korean Medicine, Gachon University, 1342 Seongnamdaero, Sujeong-gu, Seongnam-si, Gyeonggi-do, Korea

E-mail : eopchang@gachon.ac.kr · Tel : +82-31-750-5493

Received : 2016/12/06 · Revised : 2016/12/14 · Accepted : 2016/12/19

© The Society of Pathology in Korean Medicine, The Physiological Society of Korean Medicine

piSSN 1738-7698 eiSSN 2288-2529 <http://dx.doi.org/10.15188/kjopp.2016.12.30.6.412>

Available online at <https://kmpath.jams.or.kr>

발로 이어질 수 있을 것이라 기대한다.

연구대상 및 방법

최근 몇 년 사이 양적, 질적으로 폭발적인 성장을 하고 있는 MOOC(massive open online course)를 중심으로 공부 방법을 안내할 것이다. MOOC를 통해 Harvard, MIT, Stanford 등 세계 유수의 대학강의를 무료(certificates 발급시 유료)로 이용할 수 있으며, 학습자는 단순히 동영상만 포함한 수업자료를 제공받는 것이 아니라 시간에 맞춰 과제를 제출하고 시험을 보는 등, 실질적으로 수업에 참여하게 된다. 대표적인 MOOC 플랫폼으로 코세라(<https://www.coursera.org>), 에드엑스(<http://www.edx.org>), 유다시티(<http://udacity.com>), 퓨처런(<https://www.futurelearn.com>) 등이 있으며, 한국형 MOOC인 KMOOC(<https://www.kmooc.kr>) 역시 빠르게 발전하고 있다.

KMOOC를 제외하고는 대부분의 강의가 영어로 진행되지만, 영어자막이 함께 제공되므로 영어 듣기가 취약한 수강생이라도 어려움 없이 수업을 따라갈 수 있다. 각 플랫폼 별로 구체적인 운영 방식과 정책이 조금씩 다르므로 처음 시작시엔 플랫폼 별 이용방법을 숙지하고 익숙해질 필요가 있다.

MOOC의 강의들 중엔 몇 년째 꾸준히 반복적으로 제공되거나 상시 수강 가능한 형태로 열려있는 강의들도 있지만, 일회성으로 제공된 후 다시 열리지 않는 강의들도 있다. 특히 KMOOC의 경우 한국어 강의라는 점에서 접근성이 좋지만, 아직은 짧은 역사로 강의를 지속적 제공 여부가 불투명한 경향이 있다. 본 논문에서는 지속적으로 반복 제공되고 있거나 archive 형태로 제공되는 강의를 위주로 소개할 예정이다. 또한 경우에 따라 MOOC의 형태는 아니지만 동영상을 포함한 대학 강의자료를 제공하는 OCW(OpenCoursreWare)도 소개하도록 한다.

본 논문의 목적은 실용적인 공부방법 제안에 있으며, 관련 현황을 객관적으로 조사, 비교하고자 하는 것이 아니다. 더구나 저자의 MOOC 이용 경험에는 한계가 있으므로 주관적으로 선별된 강의를 소개될 수 밖에 없다. 따라서 독자들은 저자의 추천을 참고로 삼되 각자 본인의 공부 목적, 수준에 맞는 강의를 발굴하고 시도할 수 있을 것이다.

필요한 학습 내용을 다음의 5가지 항목으로 구분하여 소개하고자 한다.

1. 컴퓨터 프로그래밍, 2. 수학, 3. 통계학, 데이터 사이언스(data science), 기계학습(machine learning), 4. 복잡계 과학(complexity science), 5. 생물학, 의학 등 도메인 관련 지식.

각 항목 간에 중요도나 학습 순위를 정하는 것은 쉽지 않으며 이상적이긴 하지만 동시적 학습이 가장 좋을 것이다. 다만 구체적으로 개별 과목들간에는 선행되어야 하는 과목과 심화 과목이 존재할 수 있다.

결 과

1. 컴퓨터 프로그래밍

직접 코딩(coding)을 하지 않고서는 본격적인 연구를 할 수 없다. 마우스가 아니라 자판으로 컴퓨터와 소통하는 능력을 갖추는 것은, 문맹인이 글을 자유자재로 읽고 쓰는 능력을 얻게 되는 것에 비견될 만 하다. 코딩 능력을 갖추는 것은 실질적인 연구를 위한 시작이라고 할 수 있다. 프로그래밍에 입문하는 사람은 수많은 프로그래밍 언어 중에 어떤 것을 익혀야 하는지 많은 고민을 하게 된다. 우선적으로 다음의 언어를 추천한다.

1) 파이썬(Python)

파이썬만으로 충분할 수도 있고, 필요에 따라 다른 언어를 추가적으로 익힐 수도 있다. 파이썬은 현재 가장 많이 이용되는 프로그래밍 언어 중 하나로, 오픈소스(무료)이며 배우고 쓰기가 매우 쉽다. 보다 인간의 언어에 가깝고 컴파일(compile)없이 코드를 실행할 수 있는 인터프리트 방식이기 때문에 아이디어를 빠르게 코딩하고 결과를 확인, 수정할 수 있다. 따라서 생산성이 매우 높다. 다른 프로그래머들이 배포한 수많은 라이브러리(재사용할 수 있는 코드 모음)를 활용함으로써 코딩 시간과 번거로움을 줄일 수 있는 것도 큰 장점이다. 사용이 쉽지만 범용 언어(general-purpose language)이기 때문에 하드웨어 제어 등 시스템/임베디드 분야만 아니라면 원하는 모든 일을 할 수 있다고 해도 과언이 아니다. 또한 필요시 다른 언어로 작성된 코드와 결합하는 능력도 탁월하여, 파이썬만으로 아쉬운 부분은 파이썬 코드 내에서 다른 언어로 작성된 코드를 돌릴 수도 있다. 당장 MOOC에 개설된 컴퓨터 프로그래밍 관련 강의의 대부분이 파이썬을 채택하고 있다는 사실만으로도 이 언어의 실용성과 인기를 짐작할 수 있다.

(1) 추천강의

① Learn to Program: The Fundamentals

플랫폼: 코세라(Coursera)

제공: University of Toronto

저자 코멘트: 프로그래밍에 경험이 전혀 없는 초보자도 충분히 따라갈 수 있는 난이도. 7주 과정으로 간결하면서도 컴팩트하게 핵심적인 내용을 다루고 있다. 무엇보다도 Python Visualizer라는 시각화 프로그램을 활용하여 프로그램 실행시 파이썬 내부에서 메모리주소를 할당하고 변수가 값을 참조하는 원리를 확인시켜주는 강의 방식이 훌륭하다.

② Introduction to Computer Science and Programming Using Python

플랫폼: 에드엑스(EdX)

제공: MIT

저자 코멘트: MIT의 컴퓨터 과학 개론 수업을 그대로 옮겨온 강의. 사전 지식이 없는 초심자를 대상으로 하나 재귀함수와 같은 심화 내용을 일부 다루며 매주 부여되는 프로그래밍 과제의 난이도가 높아 상당한 집중력과 끈기를 요구한다.

③ Python for Everybody (Specialization)

플랫폼: 코세라(Courera)

제공: University of Michigan

저자 코멘트: Specialization 과정으로서 프로젝트 위주의 캡스톤(capstone) 코스를 포함하여 총 5개의 코스로 구성되어있다. 프로그래밍 입문자를 위한 Programming for Everybody로 시작

하여 Python Data Structures, Using Python to Access Web Data, Using Databases with Python 코스로 이어지며, 웹과 데이터베이스를 이용하여 인포메틱스(informatics)의 실무적인 지식과 기술을 배우기에 충분하다. 설명도 매우 친절하다.

④ Applied Data Science with Python (Specialization)

플랫폼: 코세라(Courera)

제공: University of Michigan

저자 코멘트: 역시 Specialization 과정으로서 프로젝트 위주의 캡스톤(capstone) 코스를 포함하여 총 5개의 코스로 구성되어 있다. 데이터 사이언스에 파이썬을 이용하기 위해 필요한 통계, 기계학습, 네트워크 분석 등의 지식과 라이브러리 사용법 등을 전방위적으로 다루고 있으며 이론보단 실제 적용을 위한 테크닉에 방점을 두고 있다. 위의 다른 강의들과 달리 파이썬 프로그래밍의 기초가 있는 수강자를 대상으로 한다.

2) R

데이터 사이언스를 위한 프로그래밍 언어를 추천한다면 파이썬과 함께 R 을 추천할 수 있다. 파이썬과 마찬가지로 오픈소스이며 인터프리터 언어(소스코드의 컴파일 과정 없이 한줄 한줄 바로 실행되는 언어)로서 배우기가 쉽다. 통계 언어로서 데이터 처리와 분석에 최적화되어 있으나 파이썬과 같은 범용성을 가진 언어는 아니다. 기본적으로 설치되는 통계 패키지뿐 아니라 오픈소스 언어답게 수많은 강력한 패키지를 손쉽게 공유하고 활용할 수 있다. 구체적인 연구분야와 방향에 따라 다르겠지만 인포메틱스(informatics)나 데이터 사이언스 관점에서 연구하는 시스템 생물학 연구자들은 많은 경우 파이썬과 R을 모두 활용하고 있다. 물론 파이썬만을 이용하거나 R만을 이용하는 연구자들도 많이 있으며 사실상 한가지만으로도 큰 제약은 없다.

R은 통계 언어이기 때문에 일반적으로 MOOC의 통계학이나 데이터 사이언스 강의에서 함께 다루지고 있으므로 해당 항목에서 관련 강의를 추천한다.

3) 매트랩(MATLAB)

매트랩(MATLAB)은 수학 계산, 모델링, 시뮬레이션, 신호처리 등 다양한 과학, 공학 분야에 이용되는 수치해석 프로그래밍 언어이다. 오픈 소스인 파이썬, R과 달리 Mathworks사에서 개발하여 판매하는 유료 소프트웨어이다. 학계의 연구자들은 소속 학교나 연구소가 라이선스를 구입한 경우 쉽게 이용할 수 있다. 과학, 공학 기술용 계산을 위한 전문 언어답게 다양하고 강력한 성능의 툴박스(toolbox)들이 내장되어 있으며 오픈 소스 언어와 비슷하게 이용자들이 만든 함수파일(m-file)들이 공유되기도 한다. 최근 파이썬의 NumPy, SciPy 등 매트랩의 기능을 대체하는 라이브러리들이 많이 발전함에 따라 이용자가 다소 줄어들고 있는 경향이 있으나, 여전히 많은 분야가 매트랩을 익혀야만 접근 가능한 것도 사실이다. 특히 신경과학 분야의 신호 및 영상처리(뇌영상, 전기생리학적 신호 등), 시뮬레이션 등과 같은 분야는 전통적으로 매트랩 의존도가 매우 높아 관련 연구자라면 매트랩을 익히는 것을 추천한다.

(1) 추천 강의

① Introduction to Programming with MATLAB

플랫폼: 코세라(Coursera)

제공: Vanderbilt University

저자 코멘트: 컴퓨터 프로그래밍이나 수학에 기초가 없는 사람들도 수강할 수 있는 입문 강의이며 내용이나 구성이 모두 훌륭하다.

2. 수학

구체적인 연구 분야와 형태에 따라 필요한 수학의 깊이와 범위가 달라진다. 대략적으로 구분한다면 복잡한 데이터를 이용, 데이터 사이언스 적인 접근을 위주로 하는 경우와 보다 이론적인 모델링 및 시뮬레이션을 위주로 하는 경우로 나눌 수 있을 것이다. 전자는 선형대수(linear algebra), 후자는 미적분학(calculus) 및 미분방정식(differential equation)에 대한 이해가 우선적으로 필요하다. 물론 두 접근이 상호배타적인 관계는 아니지만 연구자의 우선적인 관심 분야에 따라 학습의 우선순위를 달리 할 수 있다.

1) 선형대수학(linear algebra)

(1) 추천 강의

① Linear Algebra (by Gilbert Strang)

플랫폼: MIT OpenCourseWare(<https://ocw.mit.edu/course/s/mathematics/18-06-linear-algebra-spring-2010/>)

제공: MIT

저자 코멘트: MOOC가 아닌 MIT OpenCourseWare에서 제공하는 강의이다. 전세계적으로 호평받은 Gilbert Strang 교수의 강의 동영상 및 자료들을 이용할 수 있다. 교재를 함께 공부해야 하는데 동영상 강의보다 교재에 더 자세한 내용과 문제들이 다뤄지고 있다. 강의는 직관적인 이해를 매우 중요시하며, 고차원의 데이터(high dimensional data)를 다루기 위해 필요한 다차원 공간(high dimensional space)에 대한 수학적 직관을 마스터할 수 있도록 도와준다. 직관적 이해를 위한 설명이 때로는 지나치게 다양하고 긴 예시로 이어져 초심자에게 어렵다는 비판도 있으나, 이해가 안되는 부분은 일단 넘기고 문제를 풀며 진도를 진행한 후 다시 복습한다면 깊은 의미를 이해할 수 있다.

2) 미적분학(calculus)

한국 고등학교 미적분학 수준을 포함하여 아주 기초적인 수준부터 제공되고 있으므로 각자의 수준에 맞는 강의를 수강하면 된다.

(1) 추천 강의

① Calculus 1A, Calculus 1B, Calculus 1C

플랫폼: 에드엑스(EdX)

제공: MIT

저자 코멘트: 세 강의는 각각 differentiation(미분), integration(적분), coordinate system(극좌표계)라는 부제를 갖고 있다. MIT의 첫번째 미적분학 강의에 해당하며 고등학교 미적분을 전혀 기억하지 못하는 학습자라도 문제 없이 따라갈 수 있을 만큼 쉬운 수준부터 차근차근 진행된다. 하지만 MIT학부 강의에 해당하는 만큼 요구하는 이해도와 문제의 난이도는 가볍지 않다. 코스가 주기적으로 오픈되진 않지만 archived 상태로도 자유롭게 이용가능하다.

② Single Variable Calculus

플랫폼: 코세라(Coursera)

제공: University of Pennsylvania

저자 코멘트: 수강자가 고등학교 수준의 기초 미적분학에 대한

학습이 되어있다고 전제하고 대학교 미적분학의 핵심 내용중 하나인 테일러 급수(Taylor series)부터 시작한다. 전 강의 내용이 애니메이션으로 구성되어있어 직관적 이해를 돕는다. 응용과 활용 측면에 보다 집중하고 있고, 아기자기한 애니메이션에서 느껴지는 느낌과 달리 내용의 수준은 상당히 높다.

3) 미분방정식(differential equation)

기전 수준에서의 모델링(mechanistic modeling)을 위해서는 미분방정식에 대한 이해가 필수적이다. 선수과목으로 미적분학을 공부해야 한다.

(1) 추천 강의

① Introduction to Differential Equations

플랫폼: 에드엑스(EdX)

제공: Boston University

저자 코멘트: 친절한 설명과 다양한 예제를 바탕으로 차근차근 진행되므로 미적분학에 대한 지식이 있다면 그리 어렵지 않게 따라갈 수 있다.

3. 통계학, 데이터사이언스(data science), 기계학습(machine learning)

MOOC에서 가장 수요가 많은 과목들로, 개설되어 있는 강의의 수도 매우 많다. 일반적으로 이 분야의 학습을 위하여 미적분학 및 선형대수학의 기초가 선행되어야 하지만, 관련 수학 지식이 부족한 학습자들을 대상으로 하는 쉬운 강의들도 많이 있다.

1) 기초 통계학

통계량(statistic)과 모수(parameter), 추출(sampling), 확률(probability)과 분포(distribution) 등에 대한 개념, 대표적인 분포의 종류와 특성, 점추정(point estimation) 및 구간 추정(interval estimation), 가설 검정(hypothesis testing)의 개념과 방법 등, 통상적으로 학부 교양 통계학 강의에서 다루지는 수준의 내용을 말한다. 연구자라면 분야를 막론하고 숙지해야 하는 내용이다. 강의 난이도에 따라 기본적인 미분과 적분 지식이 필요할 수도 있고, 그렇지 않을 수도 있다. 일반적으로 MOOC에 소개되는 기초 통계학 강의를 수강하는 데에는 미적분학의 지식이 거의 필요하지 않다.

(1) 추천 강의

① Basic Statistics

플랫폼: 코세라(Coursera)

제공: University of Amsterdam

저자 코멘트: 다양한 기초 통계학 강의들이 있고, 강의들마다 다루는 범위는 크게 차이가 나지 않지만 내용의 전달력면에서는 많은 편차가 있다. 이 강의는 저자가 접해본 기초 통계학 강의중에 전달력과 구성이 가장 좋다. 지루할 틈 없는 다소 빠른 페이스의 강의 전개와, 내용과 완벽하게 싱크를 이루는 유려한 애니메이션 구성, 유쾌하고 익살스러운 스토리텔링까지 갖추고 있어 재미와 내용 전달력 모두에서 완성도가 매우 높은 '작품'이다. 실습은 R을 이용하며, 웹상에 구현된 thirty-party tool인 R lab에서 이루어진다. 물론 R에 대한 사용법도 친절하게 설명해주고 있다.

② Inferential Statistics

플랫폼: 코세라(Coursera)

제공: University of Amsterdam

저자 코멘트: 앞서 소개한 Basic Statistics에 이어지는 강의. 통계학의 핵심인 통계적 추론을 본격적으로 다룬다. 통계적 추론까지 공부하면 일반적인 기초 통계학 공부를 마쳤다고 할 수 있다.

③ Statistics with R (Specialization)

플랫폼: 코세라(Coursera)

제공: Duke University

저자 코멘트: 캡스톤 프로젝트를 포함, 총 5개의 강의로 이루어져있는 Specialization 코스. 상당히 깔끔하고 정돈된 강의 구성으로, 내용의 전달력도 훌륭하며 회귀분석(linear regression)과 베이저안 통계(Bayesian statistics)와 같이 기초 통계학보다 심화된 내용까지 다루고 있다.

2) 수리통계학 (mathematical statistics)

기초 통계학의 내용들이 자연과학, 공학, 의학, 사회과학을 막론하고 전 분야의 연구자에게 필수적이라고 한다면, 수리통계학은 보다 전문적으로 통계학의 이론적인 측면을 파고들거나, 기계학습, 인공지능, 신호처리, 계량경제학 등 수리적으로 데이터를 다루는 연관 분야들을 공부하는데 필수적이라 할 수 있다. 수리통계학 공부를 위해서는 기초 통계학의 개념이 선행되어야 하고, 상당한 수준의 미적분학에 대한 이해도와 계산능력이 갖추어져 있어야 한다. 강의에 따라서 해석학(real analysis)이나 선형대수학(linear algebra)에 대한 지식을 요구하기도 한다.

수학, 통계학 등을 학부에서 전공하지 않은 입장에선 수리통계학 공부와 함께 실제로 손을 더럽히는 계산 연습을 열심히 해야 한다. 이 부분은 머릿속 이해도 중요하지만 지난한 연습을 통해 익숙해지는 과정이 필요하다.

(1) 추천 강의

① 수리통계학 1,2

플랫폼: KOCW

제공: 부산대학교

저자 코멘트: 한국형 OpenCourseWare인 KOCW에 공개되어 있는 1년 과정의 수리통계학 강의. 코세라(Coursera)에도 존스홉킨스 대학(Johns Hopkins University)에서 제공하는 동일한 제목의 강의를 있지만, 내용의 충실성이나 범위, 설명능력 등에서 이 강의와 비교가 되지 않는다.

3) 기계학습 (machine learning) & 데이터 사이언스(data science)

최근 가장 뜨거운 관심을 받고 있는 분야로서 다양한 분야에 활용도가 높으며, 빅데이터와 함께 성장하고 있는 시스템 생물학 연구에서도 그 중요성이 커지고 있다. 많은 수요가 있는 분야인 만큼 다양한 강의들이 개설되어 있으며 현재도 계속해서 새로운 강의들이 개설되고 있다. 실무적인 능력 배양을 목표로, 이론적인 깊이 보다는 실제 활용 측면에 집중하는 강의들이 대부분이다. 따라서 원활한 수강을 위해서 수학적 능력보다는 실제 구현을 위한 코딩 능력이 더 중요한 경우가 많다.

(1) 추천 강의

① Machine Learning

플랫폼: 코세라(Coursera)

제공: Stanford University

저자 코멘트: 코세라(Coursera)의 창립자이자 현재 기계학습

연구의 최전선에서 딥러닝(deep learning)의 구루 중 한명으로 활약중인 Andrew Ng 교수의 강의. 코세라(Coursera) 초기부터 최고의 강의로 인정받아온 강의 중 하나로, 일반적으로 기계학습 입문자에게 가장 강력하게 추천된다. 수학적 기초가 없는 개발자를 수강생으로 상정하고 아주 기초적인 내용부터 차근차근 설명이 이뤄진다. 직관적인 이해를 추구하는 설명 능력이 탁월하며, 실무적으로 기계학습 활용시에 부딪히게 되는 문제들에 대한 팁을 알려주는 부분들도 훌륭하다. 수학 지식 없이도 들을 수 있지만 편미분, 선형대수학에 대한 기초지식이 있다면 보다 수월하게 들을 수 있다. 부분적으로 수학, 수리통계 지식을 활용하지 않고 설명하는 과정에서 오히려 더욱 복잡해지는 경우도 있는데, 미적분학과 선형대수, 수리통계학에 대한 지식이 있다면Stanford 대학의 정식 강의자료(<http://cs229.stanford.edu/materials.html>)를 함께 참고해도 좋다. 실습은 Octave로 이루어지는 데 Octave는 무료 소프트웨어로 사실상 MATLAB과 동일한 언어이다. MATLAB에 익숙치 않고, 따로 익힐 계획이 없는 학습자에게는 실습 부분이 아쉽겠지만, 이론 강의만으로도 공부할 가치는 충분하다. 웹상에서 발표를 판단면 누군가 실습코드를 파이썬으로 구현한 버전도 찾을 수 있다.

② Learning From Data (Introductory Machine Learning)

플랫폼: 에드엑스(EdX)

제공: Caltech

저자 코멘트: 앞서 소개한 Andrew Ng 교수의 Machine Learning이 기본적인 개념과 원리 설명에 더해 실제 활용에 초점을 맞추는 실무 중심의 강의라면, 이 강의는 난해한 이론적 설명을 정면돌파하는, 보다 학술적인 강의이다. 현재 MOOC에 개설된 대다수의 기계학습 강의들이 실용적인 활용 능력에 초점을 맞추고 있다는 점에서 다소 특이한 강의라 할 수 있는데 Caltech에서 진행된 기계학습 강의를 그대로 옮겨온 구성(주당 권장 학습 시간이 10-20시간)이기 때문이라 할 수 있다. 강의를 진행하는 Yaser S. Abu-Mostafa 교수는 기계학습 분야의 권위자이자 Teaching Awards를 수차례 수상한 바 있는 탁월한 교육자이다. 난해하고 도전적인 개념들을 명료하게 도식화하여 설명하는 능력이 탁월하다. 수강하기 위해선 미적분학과 선형대수, 기초적인 확률통계 개념, 그리고 프로그래밍 언어 중 한가지에 이미 익숙한 상태여야 하며, 보다 쉬운 기계학습 강의를 통해 기본적인 개념을 이해하고 있는 것이 수월하다.

③ Data Science (Specialization)

플랫폼: 코세라(Coursera)

제공: Johns Hopkins University

저자 코멘트: 캡스톤(Capstone) 프로젝트를 포함하여 무려 10개의 강의로 구성된 Specialization 코스. R을 주무기로 하는 데이터 사이언티스트가 갖춰야 할 광범위한 지식과 기술을 다루고 있다. Markdown, GitHub과 같이 데이터 사이언스에 필요한 기초 도구부터 R 프로그래밍, 웹과 API 및 데이터베이스에서의 데이터 추출과 전처리 방법, 통계 분석 기법 및 기계학습 활용, 그리고 데이터 기반의 어플리케이션, R 패키지 개발에 이르기까지 실무자에게 필요한 거의 모든 기술들이 다뤄지고 있다. MOOC 초창기부터 데이터 사이언스 교육 컨텐트에 공을 들여온 Johns Hopkins

University 교수진들이 자부하는 코스로서, 다루는 범위 뿐 아니라 내용도 훌륭하다. 다만 전반적으로 교수진들의 내용 전달 능력이 뛰어난 편은 아니라 기초가 부족한 초심자는 각 과목에 해당하는 다른 강의를 함께 수강하기를 권하고 싶다.

4. 복잡계 과학 (complexity science)

복잡계(complex system)란 구성요소들간의 비선형적 상호작용(nonlinear interaction)이 시스템 수준에서의 행위를 창발(emerge)하고 외부 변화에 적응, 진화해나가는 시스템을 의미한다. 복잡계 과학이란 수학, 물리학, 컴퓨터 과학 등의 방법론을 동원하여 복잡계를 정량적으로 이해하려는 과학을 의미하며 사실 하나의 명확한 이론이나 접근법이 존재한다기보다는 여러 가지 이론과 접근 방법들의 묶음으로 이해할 수 있다. 현재 시스템 생물학 연구에서 많이 이용되고 있는 네트워크 분석 기술이 복잡계 과학이 시스템 생물학 연구에 적용되고 있는 한 예이며, 향후 양질의 생물학적 데이터의 대량생산이 가속화될수록 복잡계 과학의 활용 범위도 더욱 넓어지리라 예상된다. 따라서 보다 적극적으로 복잡계과학의 이론과 방법론을 공부하는 것도 고려해볼 만 하다.

복잡계 과학 연구의 총 본산이라 할 수 있는 Santa Fe Institute에서 운영하고 있는 복잡계 과학 전문 MOOC Complexity Explorer(<https://www.complexityexplorer.org>)의 강의들을 이용하는 것을 추천한다. 모든 강의는 무료이며, 다학제적 학문의 특성상 수학, 컴퓨터 과학 등의 기초가 없는 학습자들을 위한 배려도 잘 되어있어 부담없이 공부를 시작할 수 있다. 물론 수학과 프로그래밍을 활용한 보다 심화된 강의들도 제공되고 있다.

(1) 추천 강의

① Introduction to Complexity

플랫폼: Complexity Explorer

제공: Santa Fe Institute

저자 코멘트: 복잡계 과학의 개념과 다양한 접근방법의 개요를 이해할 수 있도록 친절하게 소개하는 강의. 수학, 컴퓨터 과학의 사전지식이 필요 없다. 복잡계 과학 전체를 아우르는 강의이므로 이 분야의 입문자에게 추천한다.

② Introduction to Dynamical Systems and Chaos

플랫폼: Complexity Explorer

제공: Santa Fe Institute

저자 코멘트: 복잡계 과학의 여러 분야 중 카오스(chaos) 현상과 동역학 시스템에 초점을 맞추고 있는 강의. 고등학교 수준의 수학만으로 카오스와 동역학 시스템의 핵심적인 개념을 이해할 수 있도록 쉬운 설명을 구사한다.

③ Nonlinear Dynamics: Mathematical and Computational Approaches

플랫폼: Complexity Explorer

제공: Santa Fe Institute

저자 코멘트: 비선형 동역학과 카오스를 수학과 컴퓨터 프로그래밍을 이용하여 보다 심도 있게 이해할 수 있는 강의. 프로그래밍은 학습자가 원하는 언어를 이용, 간단한 시뮬레이션을 수행하는 정도로 충분하며 수학은 미적분에 대한 선수 지식을 요한다.

5. 생물학, 의학 등 도메인 지식(domain knowledge) 관련

한의학 연구에 시스템 생물학의 방법론을 적용하려는 연구자로서 어떤 데이터를 이용할 것인지에 따라 해당하는 영역의 공부의 달라질 것이다. 유전체 정보를 비롯한 각종 오믹스(omics) 데이터를 염두에 두고 있다면 유전체학(genomics), 생물정보학(bioinformatics) 등에 대한 이해가 필요하며, 임상 데이터(clinical data)의 이용을 염두에 두고 있다면 시스템 의학(systems medicine), 디지털 헬스케어(digital healthcare) 등의 공부가 필요할 것이다.

1) 생물학

(1) 추천 강의

① Introduction to Biology - The Secret of Life

플랫폼: 에드엑스(EdX)

제공: MIT

저자 코멘트: 유전체학(Genomics)분야의 스타중 한명인 Eric Lander의 유명 MIT 학부 강의를 그대로 옮겨온 강의. '생물학이 이렇게나 재미있고 멋진 학문이었나' 라는 생각을 하게끔 하는 명강의이다. 생물학 기초를 복습할 필요가 있다면 이 강의를 추천한다. 단순히 교실 강의를 옮겨온 것만이 아니고 시각화 툴과 같은 발전된 학습도구로 학습자의 이해를 돕는다.

② Systems Biology and Biotechnology (Specialization)

플랫폼: 코세라(Coursera)

제공: Mount Sinai

저자 코멘트: 다학제적 시스템 생물학 연구와 관련 커리큘럼 개발을 선도하는 Icahn School of Medicine at Mount Sinai에서 제공하는 Specialization 과정. 캡스톤(capstone) 프로젝트를 포함하여 총 6개의 강의로 구성되어 있으며, 오믹스 기술(omics technology), 생물정보학(bioinformatics), 동역학 모델링(dynamical modeling), 네트워크 분석(network analysis)등 시스템 생물학에서 이용되는 실험적, 이론적 접근을 모두 망라하고 있다. 각 강의 별로 요구되는 선수 과목이 있으나, 개괄적 소개에 해당하는 첫 번째 강의 Introduction to Systems Biology는 특별한 수학적지식이나 컴퓨터 프로그래밍 능력을 요구하지 않으므로 시스템 생물학적 접근에 관심 있는 입문자들은 우선적으로 공부해 볼 것을 추천한다.

③ Cellular mechanisms of brain function

플랫폼: 에드엑스(EdX)

제공: EPFL

저자 코멘트: 신경생리학의 실험적 접근과 계산신경과학적 이론 연구 모두에서 뛰어난 역량을 보여주고 있는 스위스 EPFL에서 제공하는 강의. 현재 MOOC에서 찾을 수 있는 신경생리학 강의 중에 가장 높은 수준의 완성도를 보여준다고 생각된다. 세련되면서도 명확하고 깊이 있는 설명으로 최근의 광유전체학(optogenetics) 및 이미징 기술을 적용한 네트워크 수준 연구까지 포괄하고 있어 신경생리학의 기초부터 첨단까지 공부할 수 있다.

2) 의학

(1) 추천 강의

① Data Analytics in Health - From Basics to Business

플랫폼: 에드엑스(EdX)

제공: KU Leuven

저자 코멘트: 건강보험의 환자 데이터, 유전체 데이터 등 의학 적 데이터를 데이터 마이닝(data mining)하고 기계학습 기법을 적용하여 예측 모델을 만드는 실제 케이스와 이를 비즈니스 모델로 연결시키는 방법까지 소개하고 있는 강의. 짧은 케이스 위주의 강의이지만 실제 사례를 통해 각종 의학 데이터를 활용하는 연구 방법을 배울 수 있다.

고 찰

이상으로 시스템 생물학의 방법론에 입문하는 한의학 연구자를 위한 학습 과목과 추천 강의를 5가지 항목으로 나누어 소개하였다. 여기에 소개된 학습 과목들은 분야별 입문과목들에 해당하며 심화 과목들은 거의 다루지 않았다. 항목 별로 심화 지식을 얻을 수 있는 강의들이 MOOC에 상당수 개설되어 있으므로 관심과 필요에 따라 심화과정에 도전할 수 있을 것이다. 하지만 시스템 생물학의 다학제적 연구 특성상 한 명의 연구자가 모든 관련 분야의 심화지식까지 통달하는 것은 어려운 일이며, 그럴 필요도 없다. 한 두가지의 영역에 대한 전문성을 중심으로 시스템 생물학 전반에 이해도를 갖추고, 해당 전문가들과 소통할 수 있는 것을 목표로 하는 것이 바람직한 방향일 것이다.

본 논문에서 추천한 강의들은 저자의 한정된 경험에 기반하고 있으며 최선의 선택은 아님을 다시 한번 강조하고 싶다. 강의 소개나 리뷰를 참고했을 때 매우 훌륭한 강의로 예상되지만 직접 수강해보지 않은 강의를 추천할 수는 없기에 목록에서 제외된 강의가 많다. 독자들은 본 논문의 추천을 참고 삼아 스스로 좋은 강의를 발굴하고 시도해 볼 수 있을 것이다.

마지막으로 성공적인 MOOC 학습을 위한 조언을 하고자 한다.

첫째, 필기를 해야 한다.

강의에 따라 수업 내용이 잘 정리된 lecture note가 제공되는 경우도 있고, 그렇지 않은 경우도 있다. 수업에서 따로 제공되는 정리가 없다면 스스로 노트에 강의 내용을 필기해야 한다. 필기를 하지 않고 진도를 나가다 점점 내용이 어려워지면서 앞의 내용이 기억이 나지 않을 때, 쉽게 참고할 수 있는 자료가 없다면 강의 이해도가 갑자기 떨어지며 흥미를 잃게 되고 포기하기 쉽다. 반드시 노트에 필기를 하고 복습을 하며 강의를 듣기를 권한다.

둘째, 수료증(certification)이나 Specialization(EdX의 경우 Xseries 혹은 MicroMaster)을 목표로 하는 것을 권한다.

강의에 따라 난이도 편차가 있지만 전업 학생이 아닌 입장에서 꾸준히 일정 시간을 온라인 학습에 할애하고, 더구나 머리를 싸매며 고민해도 잘 풀리지 않는 난이도의 수업을 끝까지 완수하는 것은 생각보다 훨씬 어려운 일이다. 특히 어려운 프로그래밍 숙제의 경우 확실한 보상과 목표가 없다면 바쁜 일상과 피곤함 속에 포기하기 쉽다. 수료증과 같이 객관적인 보상을 최대한 이용하여 동기를 끌어올리는 것이 좋다. 코세라(Coursera)의 경우 간단한 절차만으로도 financial aid를 받을 수 있으므로 경제적으로 부담된다면 이를 이용하자.

셋째, 스터디 그룹을 만들어 함께 공부하는 것을 권한다.

스터디 역시 수료증과 마찬가지로 포기하지 않고 끝까지 공부할 수 있도록 하는 장치가 된다. 뿐만 아니라 서로 말과 그림으로 설명하고, 질문하고, 대답하는 과정에서 얻는 학습 효과도 크다.

결 론

한의학 연구에 있어 시스템 생물학적 접근의 유용성과 가능성이 많은 주목을 받고 있음에도 불구하고, 컴퓨터, 수학, 통계 등의 관련 지식이 부족한 한의학 연구자들은 방법론 학습의 난해함으로 표면적인 관심에 머무는 경우가 많았다.

본 논문을 계기로 보다 많은 한의학 연구자들이 시스템 생물학적 연구의 실제적 역량을 갖춘 전문가로 성장해나갈 수 있기를 바라며, 나아가 한의학계의 체계적 교육 커리큘럼의 개발도 가능할 것이라 기대한다.

References

1. Westerhoff, H.V., Palsson, B.O. The evolution of molecular biology into systems biology. *Nat Biotechnol.* 22(10):1249-1252, 2004.
2. van der Greef J. Perspective: All systems go. *Nature* 480(7378):S87, 2011.
3. Wang, L., Zhou, G.B., Liu, P., Song, J.H., Liang, Y., Yan, X.J., et al. Dissection of mechanisms of Chinese medicinal formula Realgar-Indigo naturalis as an effective treatment for promyelocytic leukemia. *Proc Natl Acad Sci U S A* 105(12):4826-4831, 2008.
4. Li, S., Zhang, B., Jiang, D., Wei, Y., Zhang, N. Herb network construction and co-module analysis for uncovering the combination rule of traditional Chinese herbal formulae. *BMC Bioinformatics* 11 Suppl 11, S6, 2010.
5. Ma, T., Tan, C., Zhang, H., Wang, M., Ding, W., Li, S. Bridging the gap between traditional Chinese medicine and systems biology: the connection of Cold Syndrome and NEI network. *Mol Biosyst.* 6(4):613-619, 2010.