2018 가을학기

소셜 네트워크 데이터마이닝과 분석

담당교수 이준환
email: joonhwan@snu.ac.kr
cell: 010-9212-4975
office: 64동 405호

시간 월요일 9:00-12:50

장소 83-601

TA 이규호 (art.and.play.7@gmail.com)

수업 개요 소셜 컴퓨팅(social computing)과 라지데이터 분석(large data analysis) 등이 커뮤니케이션 분야에서도 중요한 이슈로 부상함에 따라 컴퓨터공학을 전공하지 않은 연구자들도 소셜 네트웍 시스템의 기술적, 구조적 특성을 이해할 필요가 있다. 이 수업에서는 스크립팅 프로그래밍 언어인 파이썬(Python)을 사용하여 컴퓨터 프로그래밍의 기초를 학습하고, 웹 기반 기술(web technology), 데이터베이스 등의 관련 기술에 대한 학습을 통해 실제로 소셜 네트워크 데이터를 수집하고 분석하는 방법을 배운다.

강의내용

1. 파이썬을 이용한 기초 프로그래밍
2. 웹 기반 기술(web technology)
3. 소셜 데이터 마이닝(social data mining)을 통한 데이터 분석

교재 (참고 서적)

- CodeCademy: Learn Python
- CodeCademy: Learn HTML & CSS
- 파이썬 라이브러리를 활용한 데이터 분석 (Python for Data Analysis)
- 기타 필요한 교재/논문 등은 강의 중 제공

수업진행 계획

1주 (9/3) Introduction to Social Computing

- What is Social Computing?
- What is Computer Programming?
- Why use Python?
- How to install Python on your computer.

2주 (9/10) Python Crash Course 1

- Python Basics
- Using Github

Topic 1: Social Computing Background 1

- Anderson (2008), The End of Theory: The Data Deluge Makes the Scientific Method Obsolete, Wired
- Wee et al. (2018), The influence of depression and personality on social networking, Computers in Human Behavior.

3주 (9/17) Python Crash Course 2

- Python Basics

Data Visualization 1

- Visualizing Data with Python

Topic 2: Social Computing Background 2

- Wellman (2001), Computer Networks as Social Networks, Science.
- Lazer et al. (2009), Computational Social Science, Science.

4주 (9/24) 추석연휴 휴강

5주 (10/1) Data Visualization 2

- Visualizing Data with Tableau

Topic 3: Socialogical Concepts

- Granovetter (1973), The Strength of Weak Ties, American Journal of Sociology.
- Granovetter (1983), The Strength of Weak Ties: A Network Theory Revisited, Sociological Theory.

6주 (10/8) Data Processing

- Data cleaning process

Topic 4: SNS & Internet 1

- Ellison, Steinfield & Lampe (2007), The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites, Journal of Computer Mediated Communication.
- Erickson & Kellogg (2000), Social translucence: an approach to designing systems that support social processes, ACM Transactions on Computer-Human Interaction.

7주 (10/15) Data Analysis Using numpy and pandas 1

- numpy와 pandas를 활용한 데이터의 분석

Topic 5: SNS & Internet 2

- Ginsberg et al. (2008). Detecting influenza epidemics using search engine query data. Nature.
- Goel et al. (2010). Predicting consumer behavior with Web search. PNAS.
- Lazer et al. (2014). The Parable of Google Flu: Traps in Big Data Analysis. Science.

8주 (10/22) Data Analysis Using numpy and pandas 2

- numpy와 pandas를 활용한 데이터의 분석

Topic 6: SNS & Internet 3

- Kraut et al. (1998), Internet paradox. A social technology that reduces social involvement and psychological well-being?, American Psychologist.
- Kraut et al. (2002), Internet Paradox Revisited, Journal of Social Issues.

9주 (10/29) Text Data Processing 1

- NLTK를 활용한 텍스트의 처리
- KoNLPy를 활용한 한글 텍스트의 처리

Topic 7: Network

- De Choudhury et al. (2010), Inferring relevant social networks from interpersonal communication, WWW.
- Adamic & Glance (2005), The Political Blogosphere and the 2004 U.S. Election: Divided They Blog, LinkKDD.

10주 (11/5) Text Data Processing 2

- Word2Vec
- LDA (Latent Dirichlet Allocation)

Topic 8: Twitter & Facebook 1

- Bollen, Mao & Zeng (2011), Twitter mood predicts the stock market, Journal of Computational Science.
- Marcus et al. (2011), Twitinfo: aggregating and visualizing microblogs for event exploration, SIGCHI.

11주 (11/12) Web Technology & Social Data Ming 1

- Crawling data from websites
- Crawling data from complex websites

Topic 9: Twitter & Facebook 2

- Quercia et al. (2011), Tweets from Justin Bieber's heart: the dynamics of the location field in user profiles, SIGCHI.
- Burke & Kraut (2014), Growing closer on facebook: changes in tie strength through social network site use, SIGCHI.

12주 (11/19) Web Technology & Social Data Ming 2

- Crawling data Social Media

Topic 7: Crowdsourcing & Human Computation

- von Ahn & Dabbish (2004), Labeling images with a computer game, SIGCHI.
- Bernstein et al. (2010), Soylent: a word processor with a crowd inside, UIST.

13주 (11/26) Statistical Data Analysis

- 통계적 추론방법의 학습
- ANOVA, Regression의 학습
- 클러스터링 방법론 학습

Topic 9: Analysis Methods: Case Studies 1

- Gilbert & Karahalios (2009), Predicting tie strength with social media, SIGCHI.
- De Choudhury, Counts & Horvitz (2013), Predicting postpartum changes in emotion and behavior via social media, SIGCHI.

14주 (12/3) Using Machine Learning APIs

- Google Cloud, MS Azure etc.

Topic 10: Analysis Methods: Case Studies 2

- Gilbert & Karahalios (2009), Widespread Worry and the Stock Market, AAAI.
- Kramer at al. (2014), Experimental evidence of massive-scale emotional contagion through social networks, PNAS.

15주 (12/10) Team Meeting

16주 (12/17) Final Presentation (Final Paper Submit)

과제 및 평가 Coding Assignment: 주어진 코딩 문제를 제출

- Assignment 1: 2 CodeCademy course complete (5+5 points)
- Assignment 2: Python coding exercise (10 points)
- Assignment 3: Data cleanup exercise (10 points)
- Assignment 4: Text data processing (10 points)
- Assignment 5: Webpage crawling (15 points)
- Assignment 6: Twitter crawling (15 points)
- Assignment 7: Data Analysis (Crawling+Text Analysis) (15 points)
- Assignment 8: Using ML APIs (15 points)

Paper Review: 매주 2~3개의 페이퍼를 읽고 A4 두페이지 이내로 요약하여 제출

- Paper Review: 2 points per paper

  * 페이퍼 리스트는 바뀔 수 있습니다. 바뀌는 경우 미리 공지합니다.

Final Project: 주제를 잡고 데이터를 수집, 분석하여 페이퍼 제출 (팀과제)

- Team Project: Data Analysis Project & Paper (70 points)
- Peer Review: 30 points

Absence Policy:

```
if 3 <= absent < 5:
    grade = "C+" #starting grade
elif absent >= 5:
    grade = "F"
```