Course Overview

HSS 510: NLP for HSS

Taegyoon Kim

Feb 28, 2024

Instruction Team

- Taegyoon Kim
  - Ph.D. in Political Science & Social Data Analytics
  - Works on digital politics, science and politics, computational social science
- Jaehong Kim
  - Ph.D. Student in Culture Technology
  - Works on morality, emotions, computational social science

Course Objective

We aim to

- First, **understand** key NLP techniques
  - Understanding of the mathematics/statistics of NLP models
  - Principles of how these should be used to answer research questions
  - E.g., Word embeddings
    - Approaches to quantify words based on their context
    - “car” ([1.3, …, -0.34]) ≈ “automobile” ([1.5, …, -0.24])
    - How “immigrants” are understood in history and by different groups

Course Objective

We aim to

- Second, gain insights into how models and research questions inform each other
  - Research questions → appropriate models
  - Know models → new research questions

Course Objective

We aim to

- Last, develop software proficiency in implementing NLP models for research
  - Python & R
  - AI-assisted programming
    - More than “teach how to fish”
    - Role as a capable assistant

Course Schedule

We will meet at 9am every Wednesday

- A total of 16 weeks
- **No classes** on
  - Apr 10: Election day
  - Apr 17: Mid-term period
  - May 15: Buddha's Birthday
- Student presentations: Last week (Jun 12)

Mode of Instruction

This course is in-person, except

- Week 6 (Apr 3): instructor conference
  - Zoom **individually**
- Week 10 (May 1): guest lecture (NLP in Korean)
  - Byungjoon Kim, Ph.D. in Data Science
  - Zoom **in-class**
- **Zoom meeting room link**
  - Keep your camera on please
  - Blur the background if you want

Key Tasks

Let’s look into the syllabus (also check late submission policy)

- Attendance (p.2)
- Application review discussion (p.2)
- Methods tutorial (p.3)
- Research paper (p.3)

Major Themes

- Week 2 (Mar 6): Selecting and cleaning texts
- Week 3 (Mar 13): Representing and comparing texts
- Week 4 (Mar 20): Keyword-based methods
- Week 5 (Mar 27): Supervised learning methods I
- Week 6 (Apr 3): Supervised learning methods II
- Week 9 (Apr 24): Embeddings
- Week 10 (May 1): NLP in Korean
- Week 11 (May 8): Topic models
- Week 13 (May 22): Neural NLP I
- Week 14 (May 29): Neural NLP II
- Week 15 (Jun 5): Promises and pitfalls of LLMs
- Week 16 (Jun 12): Presentations

Class Components

- Lecture + guided coding (110min–)
  - Materials will be uploaded prior to each class
- Break (10min)
- Application review discussion (20min–)
- Tutorial presentation (15min–)

Readings

Make sure that you complete assigned readings before class

- The two main textbooks
  - Grimmer, J., Roberts, M.E. and Stewart, B.M., 2022. Text as data: A new framework for machine learning and the social sciences. Princeton University Press. (can be purchased at **Kyobo**)
  - Jurafsky, D. and Martin, J.H., Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (**Freely available online**).
    - **We're using the 2023 version!**

Readings

Make sure that you complete assigned readings before class

- In the syllabus, the articles for **Application review discussion** are indicated with a †
- Required readings are required
  - Blend of 1) theoretical discussions, 2) practical guidance in applications, 3) applied research
- Optional readings are optional

A fews recommendations

Please

- Prepare texts of your interest as early as possible
- You will **never** have as much time for learning new things as you do now
- Do **not** focus on just one or two techniques that suit your immediate research interests
    - Knowing techniques will allows you to discover new research topics

Logistics for next week

Please

- Complete **pre-course survey**
  - Submit by noon Mar 5 (Tue)
  - Lectures, guided coding, and readings will reflect different levels of comprehensions and demands
- We will discuss selecting texts
  - While you do the readings, think (hard) about 1) what texts would be necessary to tackle your questions and/or 2) what questions can be answered with your texts

Next week's class

- Self-introduction
- Lecture
- Guided coding
- (No **Application review discussion** nor **Methods tutorial**)