

School of Computer Science, University of Guelph

CIS*6190/DATA*6400 (Winter 2022)

Machine Learning for Sequential Data Processing [0.50]

Instructor: *Dr. Fei Song*

Email: fsong@uoguelph.ca

Class Hours: Mondays, 2:30 – 3:50 pm (Zoom Meetings in Reynolds 1101)

Wednesdays, 2:30 – 3:50 pm (Virtual Zoom Meetings)

Office Hours: Tuesdays and Fridays, 3:30 – 5:00 pm (Zoom Meetings)

Course Website: <https://courselink.uoguelph.ca/>

Graduate Teaching Assistant: Bazyli Debowski

Email: bdebowsk@uoguelph.ca

Office Hours: Thursdays, 3:00 – 4:00 pm (Zoom Meetings: starting in the week of Jan. 17 and ending in the week of Apr. 4)

Overview

Machine Learning is well-suited for the automatic analysis of Sequential Data such as Natural Language Text, Time Series, and Biological Data. For Winter 2022, we will be mostly focusing on Natural Language Processing, which is an interdisciplinary area among Mathematics, Information Theory, Linguistics, and Computer Science. It has been applied successfully to a wide range of problems such as Speech Recognition, Information Retrieval and Extraction, Text Classification, Sentiment Analysis, Automatic Summarization, and Machine Translation. Due to the sequential nature of language data, similar techniques have also been applied to the analysis of Time Series and Biological Data. This course provides an introduction to this emerging field, with emphasis on the applications of machine learning techniques. The students should have some experience in Python programming, as well as good understanding of the related mathematics, including Linear Algebra, Probability, and Statistics. In addition to attending lectures, the students are required to review the current literature and present two papers in the class. They are also required to complete two warm-up assignments and an implementation project that applies the related technique(s) for real-world applications. By the end of the semester, the students can not only gain good understanding for major and recent machine learning techniques and many related applications for Natural Language Processing, but also build solid development skills for data preprocessing and project implementation to solve real-world problems.

Listed in the following is a set of topics that we intend to cover in the course:

- Introduction to Natural Language Processing
- Language Modeling and N-gram Models
- Information Retrieval Models and Implementations
- Machine Learning for Text Classification and Sentiment Analysis
- Introduction to Deep Learning Models
- Static and contextualized word embeddings
- Unsupervised Machine Learning Methods

Evaluation Scheme

- Warmup Assignments: 2 x 15% (due on Feb. 2 and Feb. 18, respectively)
- Research Presentations: 2 x 10% (scheduled during weeks 5 to 12)
- Term Project: 50% (topic selections by Mar. 11, implementations by Apr. 20)

Late submissions are not encouraged, and there will be a deduction of marks by 10% for one day late, 25% for two days late, and 50% for three days late. No marks will be given for late submissions that are more than three days late.

Recommended References

Chris Manning and Hinrich Schütz. *Foundations of Statistical Natural Language Processing*. The MIT Press, 1999.

Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Francis Bach. *Deep Learning*. The MIT Press, 2016.

Steven Bird, Ewan Klein, and Edward Loper. *Natural Language Processing with Python*. O'Reilly, 2009.

Sebastian Raschka and Vahid Mirjalili. *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow*. Second Edition. Packt Publishing, 2017.

Calendar Description:

This course emphasizes machine learning for sequential data processing. It covers common challenges and pre-processing techniques for sequential data such as text, biological sequences, and time series data. Students are exposed to machine learning techniques, including classical methods and more recent deep learning models, so that they obtain the background and skills needed to confront real-world applications of sequential data processing. May be offered in conjunction with [DATA*6400](#).

Restriction(s): Credit may be obtained for only one of [CIS*6190](#) or [DATA*6400](#)

Email Communication

As per university regulations, all students are required to check their <uoguelph.ca> e-mail account regularly: e-mail is the official route of communication between the University and its students.

When You Cannot Meet a Course Requirement

When you find yourself unable to meet an in-course requirement because of illness or compassionate reasons, please advise the course instructor (or designated person, such as a teaching assistant) in writing, with your name, id#, and e-mail contact. See the Undergraduate Calendar for information on regulations and procedures for Academic Consideration (<https://calendar.uoguelph.ca/undergraduate-calendar/undergraduate-degree-regulations-procedures/academic-consideration-appeals-petitions/>).

Drop Date

Courses that are one semester long must be dropped by the end of the last day of classes; two-semester courses must be dropped by the last day of classes in the second semester. The regulations and procedures for dropping courses are available in the Undergraduate Calendar (<https://calendar.uoguelph.ca/undergraduate-calendar/undergraduate-degree-regulations-procedures/dropping-courses/>).

Copies of Out-Of-Class Assignments

Keep paper and/or other reliable back-up copies of all out-of-class assignments: you may be asked to resubmit work at any time.

Accessibility

The University promotes the full participation of students who experience disabilities in their academic programs. To that end, the provision of academic accommodation is a shared responsibility between the University and the student.

When accommodations are needed, the student is required to first register with Student Accessibility Services (SAS). Documentation to substantiate the existence of a disability is required, however, interim accommodations may be possible while that process is underway.

Accommodations are available for both permanent and temporary disabilities. It should be noted that common illnesses such as a cold or the flu do not constitute a disability.

Use of the SAS Exam Centre requires students to make a booking at least 14 days in advance, and no later than November 1 (fall), March 1 (winter) or July 1 (summer). Similarly, new or changed accommodations for online quizzes, tests and exams must be approved at least a week ahead of time. More information: www.uoguelph.ca/sas

Academic Misconduct

The University of Guelph is committed to upholding the highest standards of academic integrity and it is the responsibility of all members of the University community – faculty, staff, and students – to be aware of what constitutes academic misconduct and to do as much as possible to prevent academic offences from occurring. University of Guelph students have the responsibility of abiding by the University's policy on academic misconduct regardless of their location of study; faculty, staff and students have the responsibility of supporting an environment that discourages misconduct. Students need to remain aware that instructors have access to and the right to use electronic and other means of detection.

Please note: Whether or not a student intended to commit academic misconduct is not relevant for a finding of guilt. Hurried or careless submission of assignments does not excuse students from responsibility for verifying the academic integrity of their work before submitting it. Students who are in any doubt as to whether an action on their part could be construed as an academic offence should consult with a faculty member or faculty advisor.

The Academic Misconduct Policy is outlined in the Undergraduate Calendar (<https://calendar.uoguelph.ca/undergraduate-calendar/undergraduate-degree-regulations-procedures/academic-misconduct/>).

Recording of Materials

Presentations which are made in relation to course work—including lectures—cannot be recorded or copied without the permission of the presenter, whether the instructor, a classmate or

guest lecturer. Material recorded with permission is restricted to use for that course unless further permission is granted.

Resources

The Academic Calendars (<https://www.uoguelph.ca/registrar/calendars>) are the source of information about the University of Guelph's procedures, policies and regulations which apply to undergraduate, graduate and diploma programs.

Disclaimer

Please note that the ongoing COVID-19 pandemic may necessitate a revision of the format of course offerings, changes in classroom protocols, and academic schedules. Any such changes will be announced via Courselink and/or class email.

This includes on-campus scheduling during the semester, mid-terms and final examination schedules. All University-wide decisions will be posted on the COVID-19 website (<https://news.uoguelph.ca/2019-novel-coronavirus-information/>) and circulated by email.

Illness

Medical notes will not normally be required for singular instances of academic consideration, although students may be required to provide supporting documentation for multiple missed assessments or when involving a large part of a course (e.g., final exam or major assignment).

COVID-19 Safety Protocols

For information on current safety protocols, follow these links:

- <https://news.uoguelph.ca/return-to-campus/how-u-of-g-is-preparing-for-your-safe-return/>
- <https://news.uoguelph.ca/return-to-campus/spaces/#ClassroomSpaces>

Please note, that these guidelines may be updated as required in response to evolving University, Public Health or government directives.