MULTIMODAL MACHINE LEARNING – FALL 2020

Description | Schedule | Grades | Bibliography

Instructor: Prof. Louis-Philippe Morency, morency@cs.cmu.edu

Teaching assistants:
- Prakhar Gupta prakharg@cmu.edu
- Paul Liang, pliang@andrew.cmu.edu
- Martin Ma qianlim@cmu.edu
- Shikib Mehri amehri@andrew.cmu.edu
- Torsten Wörtwein twoertwe@cs.cmu.edu

Time: Tuesdays and Thursday, 3:20pm-4:40pm

Classrooms: Remote teaching – Zoom (see links in CMU Canvas)

Recommended preparation: This is a graduate course designed primarily for PhD and research master students at LTI, MLD, CSD, HCII and RI; others, for example (undergraduate) students of CS or from professional master programs, are advised to seek prior permission of the instructor. It is required for students to have taken an introduction machine learning course such as 10-401, 10-601, 10-701, 11-663, 11-441, 11-641 or 11-741. Prior knowledge of deep learning is recommended. Students should have proper academic background in probability, statistic, and linear algebra. Programming knowledge in Python is also strongly recommended.

Introduction and Purposes

Multimodal machine learning (MMML) is a vibrant multi-disciplinary research field which addresses some of the original goals of artificial intelligence by integrating and modeling multiple communicative modalities, including linguistic, acoustic, and visual messages. With the initial research on audio-visual speech recognition and more recently with language & vision projects such as image and video captioning, this research field brings some unique challenges for multimodal researchers given the heterogeneity of the data and the contingency often found between modalities. This course will teach fundamental mathematical concepts related to MMML including multimodal alignment and fusion, heterogeneous representation learning and multi-stream temporal modeling. We will also review recent papers describing state-of-the-art probabilistic models and computational algorithms for MMML and discuss the current and upcoming challenges.

The course will present the fundamental mathematical concepts in machine learning and deep learning relevant to the five main challenges in multimodal machine learning: (1) multimodal representation learning, (2) translation & mapping, (3) modality alignment, (4) multimodal fusion and (5) co-learning. These include, but not limited to, multimodal auto-encoder, deep canonical correlation analysis, multi-kernel learning, attention models and multimodal recurrent neural networks. The course will also discuss many of the recent applications of MMML including multimodal affect recognition, image and video captioning and cross-modal multimedia retrieval.

Course format

Lectures will be performed Tuesdays and Thursdays at 3:20pm. Each lecture will focus on a specific mathematical concept related to multimodal machine learning. These lectures will be given by the course instructor, a guest lecturer or a TA.

Canvas We will use CMU Canvas as a central hub for the course. From Canvas, you can access the links to the live lectures (using Zoom). You can also connect to our course discussion platform Piazza (see more details below) and the grading platform Gradescope directly from Canvas. Quizzes will be done using Canvas as well.

Zoom We will use the Zoom video platform for the live lectures on Tuesdays and Thursdays. Links to the live lectures will be available on Canvas. The lectures will also be recorded to allow students to watch them again later. Please make sure that your Internet connection and equipment are set up to use Zoom. During our class meetings, please keep your mic muted. If you have a question or want to answer a question, please use the chat or the "raise hand" feature (available when the participant list is pulled up). A TA will be monitoring these channels in order to share this information with the instructor. All course lectures. will be audio/video recorded, to allow students to watch asynchronously the lectures, if needed.

Piazza We will be using Piazza for class communication and announcement. The system is highly catered to getting you help fast and efficiently from classmates, the TAs and the instructor. Rather than emailing questions to the teaching staff, you are encouraged to post your questions on Piazza. You can post privately to the instructor and TAs through Piazza website. Piazza can be accessed from the course Canvas page, or directly at this URL:

https://piazza.com/cmu/fall2020/11777/home

Gradescope Students are asked to submit their project assignments through the website Gradescope. This platform will be used for grading and to handle any request for re-grading. Gradescope can be access from the course Canvas page.

Course Material

Required:

- Reading material will be based on published technical papers available via the ACM/IEEE/Springer digital libraries or freely available online (e.g., arxiv.org). All CMU students have already free access to these digital archives.
- For project assignments, previous experience in Python programming is expected

Optional:

- *Deep Learning*, Ian Goodfellow, Yoshua Bengio and Aaron Courville, MIT Press, 2016 (freely available at http://www.deeplearningbook.org)
- *The Handbook of Multimodal-Multisensor Interfaces*, Sharon Oviatt, Bjoern Schuller, Philip R. Cohen, Daniel Sonntag, Gerasimos Potamianos and Antonio Kruger, Volumes 1, 2 and 3, 2017-2019 (available through CMU Library online)
- *Machine Learning for Audio, Image and Video Analysis: Theory and Applications*, Francesco Camastra and Alessandro Vinciarelli, Springer, 2008, DOI: 10.1007/978-1-84800-007-0 (freely available on SpringerLink for CMU students)

- *Multimodal Processing and Interaction,* Gros, Potamianos and Maragos, SpringerLink, 2008, DOI: 10.1007/978-0-387-76316-3 (freely available on SpringerLink for CMU students)
- *Multimodal Signal Processing: Theory and applications for human-computer interaction* by Jean-Philippe Thiran, Ferran Marqués and Hervé Bourlard. Academic Press, ISBN: 978-0-12-374825-6

Course Topics

*** Exact topics may change based on student interests and time restrictions. ***

Classes	Tuesday Lectures	Thursday Lectures
Week 1 9/1 & 9/3	**Course introduction** • Research and technical challenges • Course syllabus and requirements	**Multimodal applications and datasets** • Research tasks and datasets • Team projects
Week 2 9/8 & 9/10	**Basic concepts: neural networks** • Language, visual and acoustic • Loss functions and neural networks	**Basic concepts: network optimization** • Gradients and backpropagation • Practical deep model optimization
Week 3 9/15 & 9/17	**Visual unimodal representations** • Convolutional kernels and CNNs • Residual network and skip connection	**Language unimodal representations** • Gated networks and LSTM • Backpropagation Through Time
Week 4 9/22 & 9/24	**Multimodal representation learning** • Multimodal auto-encoders • Multimodal joint representations	**Coordinated representations** • Deep canonical correlation analysis • Non-negative matrix factorization
Week 5 9/29 & 10/1	**Multimodal alignment** • Explicit - dynamic time warping • Implicit - attention models	**Alignment and representation** • Self-attention models • Multimodal transformers
Week 6 10/6 & 10/8	***First project assignment*** *(live working sessions instead of lectures)*	
Week 7 10/13 & 10/15	**Alignment and translation** • Module networks • Tree-based and stack models	**Probabilistic graphical models** • Dynamic Bayesian networks • Coupled and factor HMMs
Week 8 10/20 & 10/22	**Discriminative graphical models** • Conditional random fields • Continuous and fully-connected CRFs	**Neural Generative Models** • Variational auto-encoder • Generative adversarial networks
Week 9 10/27 & 10/29	**Reinforcement learning** • Markov decision process • Q learning and policy gradients	**Multimodal RL** • Deep Q learning • Multimodal applications
Week 10 11/3 & 11/5	**Fusion and co-learning** • Multi-kernel learning and fusion • Few shot learning and co-learning	**New research directions** • Recent approaches in multimodal ML
Week 11 11/10 & 11/12	***Mid-term project assignment*** *(live working sessions instead of lectures)*	
Week 12 11/17 & 11/19	**Embodied Language Grounding** • Connecting Language to Action • Guest lecture: Yonatan Bisk	**Multimodal language acquisition** • Learning from multimodal data • Guest lecture: Graham Neubig
Week 13 11/24 & 11/26	***Thanksgiving week*** *(no lectures)*	
Week 14 12/1 & 12/3	**Learning to connect text and images** • Discourse approaches, text & images • Guest lecture: Malihe Alikhani	**Bias and fairness** • Computational ethics • Guest lecture: Yulia Tsvetkov
Week 15 12/8 & 12/10	***Final project assignment*** *(live working sessions instead of lectures)*	

Project Assignments and Timeline

(See Piazza for additional information)

- **Dataset preferences** (Due on Tuesday 9/8 at 8pm ET) – Let us know your preferences for the datasets that you would like to use for the course project. This will help with the team matching process.
- **Project Pre-proposal** (Due on Wednesday 9/16 at 8pm ET) – You should have selected your teammates, dataset, and task. Submit a 1-page pre-proposal plan.
- **First assignment** (Due Friday 10/9 at 8pm ET for the presentations and due Sunday 10/11 at 8pm ET for the reports) – This assignment focuses on unimodal representations and requires a good literature review on the topic of your proposed project
    - **Peer feedback** (Due before Sunday 10/18 8pm ET) – Students are asked to watch other proposal presentations and share constructive feedback.
- **Midterm assignment** (Due Friday 11/13 at 8pm ET for the presentations and due Sunday 11/15 at 8pm ET for the reports) – Students are asked to implement and evaluate state-of-the-art baseline models on their project dataset.
    - **Peer feedback** (Due before Sunday 11/22 8pm ET) – Students are asked to watch other midterm presentations and share constructive feedback.
- **Final assignment** (Due Friday 12/11 at 8pm ET for the presentations and due Sunday 12/13 at 8pm ET for the reports) – Students should explore new ideas to model their multimodal research project.

Grades

Remember: If you registered for this class, you have until November 9th to change your grade in this course from a letter grade to a Pass/Fail grade.

- **Grading breakdown**
    - Lecture participation and highlights 16%
    - Reading assignments 16%
    - Course project assignments
        - Project preferences and pre-proposal 3%
        - First project assignment and presentation 15%
        - Mid-term project assignment and presentation 20%
        - Final project assignment and presentation 30%
- **Lecture participation and highlights**
    - Lectures can be attended live (using Zoom) or watched later. Students are encouraged to attend lectures live as often as possible, to allow them to ask live clarification questions, if needed. Some lectures will also contain some live survey questions.
    - While watching the lecture (either live or recorded video), students are required to fill a form where they include their main takeaways from the lecture (aka, highlights).
    - The form should be submitted within 42 hours from the scheduled end of the live lecture. For example, if the lecture was scheduled to end at 4:40pm ET on Tuesday, then the highlight form is due Thursday at 10:40am ET.

- Students need to use the provided online template for the highlight form. This form was designed for two main purposes: (1) help students for taking active notes during lectures, and (2) offer students the opportunity to ask questions about the content of the lectures.
    - The lecture is split in three 30-minute segments (the last segment may be shorter).
    - For each segment, students are asked to include a short statement summarizing the main points of the past segment.
    - Students are also asked to include their main take-away messages for the whole lecture.
    - Optionally students can also write down a question (with corresponding slide number) related to the segment.
- The student's questions will be reviewed by TAs and instructor. The most popular questions will be answered using Piazza, or with extra information during the following lecture. Students are always welcome to post questions directly on Piazza at any time if they would like clarifications or have a follow-up question.
- These highlight forms will not be required for the first week and for the Thanksgiving week. Also, no forms are expected for weeks when a project assignment (first, midterm or final) is due. We expect about 20 lectures where highlight forms need to be submitted.
- Each submitted form will be graded for 1.0 point. The top 16 scores will be kept for the lecture participation final grade.

- **Reading assignments**
    - Reading assignments are designed to complement the lectures and showcase recent state-of-the-art research. Most reading assignments will consist of multiple research papers, sometimes accompanied by optional readings. The list of research papers will be released at the latest on the Monday of each week.
    - To encourage exchange of ideas and knowledge between students, each student will be part of one study group. A study group consists of 9-10 students. These groups will be randomly created, to encourage diversity in these groups. Each study group will have its own discussion forum to ask questions and share ideas.
    - The reading assignments consist of two main parts: (a) submission of discussion post summarizing the paper you read that week (see more details below), and (b) active participation in the follow-up discussions, including at least 2 extra posts.
        - For each reading assignment, each student is required to read only one research paper (out of all assigned papers). Students need to write a summary statement for their paper and post it in the discussion forum before Friday 8pm ET. These summary posts will allow other study group members to learn about the papers they did not read directly, and possibly ask follow-up questions.
        - During the 7-day period of the reading assignment (Monday 8pm ET until the following Monday 8pm ET), each student is expected to post their summary and two extra posts. These extra posts can be follow-up questions, additional information, answers, or new insights. These extra posts can be about their own paper or the other papers.
    - Reading assignments will be released weekly, with exceptions when a project assignment (first, midterm or final) is due the same week. Also, no reading

assignments during the first week and during Thanksgiving week. We expect 10 reading assignments during the semester.
    - Each reading assignment is worth 2.0 points: 1.0 point for the paper summary and 1.0 point for the extra posts in the discussion. We plan to keep your top 8 reading assignment scores.
- **Course project assignments**
    - The goal of the course project is to experiment with state-of-the-art multimodal algorithms and computational models.
    - Students should create teams between 3 to 5 students preferably (special approval will be required for larger teams; no smaller teams will be allowed). The size and depth of the project should be adjusted to reflect the size of the team.
    - Each team is required to create a code repository (github) for their project. All project members should be included in this project and should actively use it. It is important that all team members participate equally to the project. The first project assignment and follow-up reports (midterm and final reports) will need to outline the tasks of each student. If any team member has concern in the participation level of other members, they should contact the instructor and/or TA as promptly as possible.
    - Students have flexibility in the selection of their project topic. The project should be directly aligned with the course content and include at the minimum two modalities, preferably language and vision. At the beginning of the semester, the instructor will propose a set of research problems and datasets which can be used for the course projects.
    - **Pre-proposal:** We ask students to prepare a pre-proposal early in the semester to them establish their research topic for the course project. The pre-proposal will also help with team formation, in the rare eventuality when students are still looking for teammates.
    - **First project assignment:** The first project assignment consists of a written report and an oral presentation (which will be performed remotely, with pre-recorded videos). This assignment has two main goals: describe in more details the plan for the course project (aka., "proposal") and perform some unimodal analyses on the multimodal dataset and problem.
        - **Peer feedback:** Following the submission of the proposal video recordings, students will be asked to watch the videos and share feedback. Each student will be assigned a subset of videos to watch. Feedback will also be given by instructor and TAs.
    - **Midterm project assignment:** The midterm project assignment is designed to implement multimodal baseline models and perform some error analysis on these results. This assignment also has two components: written report and oral presentation. By the submission time for the midterm assignment, students should have already implemented some of the state-of-the-art baseline models for their selected multimodal task and dataset.
        - **Peer feedback:** Like the first assignment, students will be asked to watch the midterm presentation videos and share feedback. Each student will be assigned a subset of videos to watch. Feedback will also be given by instructor and TAs.
    - **Final project assignment:** Using the same dataset and task selected for the midterm report, the final project assignment is designed to explore new research ideas. This assignment is not graded based on the quality of the results, but instead on the

exploration of new ideas (e.g., better accuracy results will not mean better course grade). Students are encouraged to explore new research directions. The final project assignment also contains written and oral components.

Note about late submissions

In general, submitting assignments on time lets the instructional team provide feedback in a more timely and efficient manner. Timely submissions are particularly important for assignments with discussions and peer feedback, such as the reading assignments and the project assignments. Since things happen, please contact me and the TAs as soon as possible (the best option is usually via Piazza) if you are not able to submit your assignment in time. For the reading assignments and lecture participation, late submissions (up to 48 hours) will be deducted 0.5 point. If you must submit beyond 48 hours past the due date, please contact me and the TAs as soon as possible so we can properly plan.

For the reading assignment summaries, each student gets one (1) wild card to be used as a way to extend by up to 24 hours their deadline for the summary deadline (which is usually Fridays at 8pm). No partial credit for the wild card (e.g., it is not possible to use only half of the card, with two times 12 hours). There is no need to send a note via Piazza about this wild card. We will automatically use your wild card the first time you submit your summary late.

The default rule for the project assignments will be that you are eligible for 90% of the grade the first 48 hours that the assignment is late. Each team will get two (2) wild cards, to be used with the project assignment deadlines:

- Each wild card allows the team to submit their assignment late for up to 24 extra hours.
- These wild cards can be used together (for a total of 48 hours), or separately (2 separate extensions of 24 hours).
- Each wild card can be used for any of these 6 deadlines:
    - First presentation deadline
    - First report deadline
    - Midterm Presentation deadline
    - Midterm report deadline
    - Final presentation deadline
    - Final report deadline
- No partial credits for the wild cards (e.g., you cannot use only 40% of a wild card).
- Each team needs to send a message on Piazza (private note to instructors) BEFORE the deadline to notify TAs and instructor about the intent to use a wild card (or two).

Accommodations for Students with Disabilities

If you have a disability and have an accommodations letter from the Disability Resources office, I encourage you to discuss your accommodations and needs with me as early in the semester as possible. I will work with you to ensure that accommodations are provided as appropriate. If you suspect that you may have a disability and would benefit from accommodations but are not yet registered with the Office of Disability Resources, I encourage you to contact them at access@andrew.cmu.edu.

Statement on Student Wellness

This semester is unlike any other. We are all under a lot of stress and uncertainty at this time. Attending Zoom classes all day can take its toll on our mental health. Make sure to move regularly, eat well, and reach out to your support system or me (morency@cs.cmu.edu) if you need to. We can all benefit from support in times of stress, and this semester is no exception.

As a student, you may experience a range of challenges that can interfere with learning, such as strained relationships, increased anxiety, substance use, feeling down, difficulty concentrating and/or lack of motivation. These mental health concerns or stressful events may diminish your academic performance and/or reduce your ability to participate in daily activities. CMU services are available, and treatment does work. You can learn more about confidential mental health services available on campus at: http://www.cmu.edu/counseling/. Support is always available (24/7) from Counseling and Psychological Services: 412-268-2922.

Diversity statement

Every individual must be treated with respect. The ways we are diverse are many and are fundamental to building and maintaining an equitable and an inclusive campus community. These include but are not limited to: race, color, national origin, sex, disability, age, sexual orientation, gender identity, religion, creed, ancestry, belief, veteran status, or genetic information. We at CMU, will work to promote diversity, equity and inclusion not only because it is necessary for excellence and innovation, but because it is just. Therefore, while we are imperfect, we all need to fully commit to work, both inside and outside of our classrooms to increase our commitment to build and sustain a campus community that embraces these core values.

It is the responsibility of each of us to create a safer and more inclusive environment. Incidents of bias or discrimination, whether intentional or unintentional in their occurrence, contribute to creating an unwelcoming environment for individuals and groups at the university. If you experience or observe unfair or hostile treatment on the basis of identity, we encourage you to speak out for justice and support in the moment and/or share your experience using the following resources:

- Center for Student Diversity and Inclusion: csdi@andrew.cmu.edu, (412) 268 2150, www.cmu.edu/student-diversity
- Report-It online anonymous reporting platform: www.reportit.net username: tartans password: plaid

All reports will be acknowledged, documented, and a determination will be made regarding a course of action. All experiences shared will be used to transform the campus climate to be more equitable and just.

Bibliography

Reading lists from Fall 2018 and Fall 2019 courses are available on Piazza:
https://piazza.com/cmu/fall2018/11777/resources
https://piazza.com/cmu/fall2019/11777/resources

The reading list for Fall 2020 semester will also be posted on Piazza website.