Instructor: Frank Ferraro

ITE 358
ferraro@umbc.edu
Monday: 2:15-3
Tuesday: 11:00-11:30
by appointment

Natural language processing
Semantics
Vision & language processing
Learning with low-to-no supervision
TA: Devajit Asem

Location TBD
devajit.asem@umbc.edu
Wednesday: 4-5pm
Friday: 2-3pm
by appointment

Databases
NLP
IR (information retrieval)
Web development
Q: What is NLP (natural language processing?)
Deep Learning
Natural Language Processing

What society thinks I do
What my friends think I do
What other computer scientists think I do

What mathematicians think I do
What I think I do
What I actually do

tensorflow
"from theano import *"
Artificial Intelligence Is Now a Pentagon Priority. Will Silicon Valley Help?

The Defense Department, believing that A.I. research should be a national priority, has called on the White House to “inspire a whole of country effort.”

By CADE METZ

Alexa vs. Siri vs. Google: Which Can Carry on a Conversation Best?

Digital assistants from Amazon, Apple and Google can only have meager back and forth exchanges with us. Listen to how that tells us something about where they’re going in the future.

By KEITH COLLINS and CADE METZ

Google Employees Protest Secret Work on Censored Search Engine for China

About 1,400 of the internet company’s employees have signed a letter demanding transparency, saying censored search results raise “urgent moral and ethical issues.”

By KATE CONGER and DAISUKE WAKABAYASHI
Potential Applications

ASR (automatic speech recognition)

Machine translation

Natural language generation

Document labeling/classification

Document summarization

Corpus exploration

Relation/information extraction

Entity identification
Potential Applications

Q: What’s an example?

ASR (automatic speech recognition)

- Machine translation
- Natural language generation
- Document labeling/classification
- Document summarization
- Corpus exploration
- Relation/information extraction
- Entity identification
Automatic speech recognition
Potential Applications

ASR (automatic speech recognition)

Machine translation

Natural language generation

Document labeling/classification

Document summarization

Corpus exploration

Relation/information extraction

Entity identification

Q: What’s an example?
SPORTS
Google Translate

Document classification

Machine translation

Les talibans mènent des attaques-suicides dans Kaboul

Air France s'attend à une perte historique pour l'exercice 2009-2010

Rendez-vous

Edition Abonnés: profitez de 71 dépêches thématiques
Potential Applications

ASR (automatic speech recognition)

Machine translation

Natural language generation

Document labeling/classification

Document summarization

Corpus exploration

Relation/information extraction

Entity identification

Q: What’s an example?
Hi all,
We wanted to invite you to join us for an early Thanksgiving on November 22nd, beginning around 2PM. Please bring your favorite dish! RSVP by next week.

Dave

Hi team,

The server appears to be dropping about 10% of requests (see attached dashboards). There hasn't been a new release since last night, so I'm not sure what's going on. Is anyone looking into this?

...
A massive climate change study is canceled because of climate change

By Doug Criss, CNN

(CNN) — A $17 million study of climate change in the Canadian Arctic has been nixed for now -- because of climate change.

A team of scientists from the University of Manitoba and four other schools were in the middle of the first leg of a four-year study of how climate change is affecting the areas around the Hudson Bay, the university said in statement. The study, named BaySys, started last month, and the scientists were traveling on the Canadian Research Icebreaker CCGS Amundsen.

But because of warmer temperatures in the Arctic, hazardous sea ice is traveling farther south than usual. The Amundsen, which is part of the Canadian Coast Guard fleet, has been diverted several times because its ice-breaking capabilities have been needed to help out in rescue efforts along Newfoundland’s northeast coast. All of the delays and concerns about safety forced
Course Goals

Be introduced to some of the core problems and solutions of NLP (big picture)
Course Goals

Be introduced to some of the core problems and solutions of NLP (big picture)

Learn different ways that success and progress can be measured in NLP
Course Goals

Be introduced to some of the core problems and solutions of NLP (big picture)
Learn different ways that success and progress can be measured in NLP
Relate to statistics, machine learning, and linguistics
Implement NLP programs
Course Goals

Be introduced to some of the core problems and solutions of NLP (big picture)
Learn different ways that success and progress can be measured in NLP
Relate to statistics, machine learning, and linguistics
Implement NLP programs
Read and analyze research papers
Practice your (written) communication skills
Adminstrivvia
Web Presence

WWW

Schedule, slides, assignments, readings, materials, syllabus here

https://www.csee.umbc.edu/courses/undergraduate/473/f19

Course announcements, Q&A, discussion board here

https://piazza.com/umbc/fall2019/cmsc473673
Please Read the Syllabus (On the Website)

https://www.csee.umbc.edu/courses/undergraduate/473/f19/content/materials/syllabus.pdf
## Grading

<table>
<thead>
<tr>
<th>Component</th>
<th>473</th>
<th>673</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignments</td>
<td>45%</td>
<td>30%</td>
</tr>
<tr>
<td>Midterm</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>Graduate Paper</td>
<td>---</td>
<td>30%</td>
</tr>
<tr>
<td>Course Project</td>
<td>45%</td>
<td>30%</td>
</tr>
</tbody>
</table>
Computation of Component Grades

Each component (e.g., “Assignment” component) is:
max(micro-average, macro-average)
Computation of Component Grades

Each component (e.g., “Assignment” component) is:
\[ \max(\text{micro-average}, \text{macro-average}) \]

Assignment grades (not representative)

65/90
95/100
95/110
100/110
Computation of Component Grades

Each component (e.g., “Assignment” component) is:
max(micro-average, macro-average)

\[
\text{microaverage} = \frac{65 + 95 + 95 + 100}{90 + 100 + 110 + 110} \approx 86.59\%
\]

\[
\text{macroaverage} = \frac{1}{4} \left( \frac{65}{90} + \frac{95}{100} + \frac{95}{110} + \frac{100}{110} \right) \approx 86.12\%
\]

Assignment grades (not representative)

We’ll learn what these are in the semester
Computation of Component Grades

Each component (e.g., “Assignment” component) is:

\[
\max(\text{micro-average}, \text{macro-average})
\]

\[
\begin{align*}
65/90 & \quad \text{microaverage} = \frac{65 + 95 + 95 + 100}{90 + 100 + 110 + 110} \approx 86.59\% \\
95/100 & \quad \text{macroaverage} = \frac{1}{4} \left( \frac{65}{90} + \frac{95}{100} + \frac{95}{110} + \frac{100}{110} \right) \approx 86.12\%
\end{align*}
\]
## Final Grades

<table>
<thead>
<tr>
<th>≥</th>
<th>Letter</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
<td>A</td>
</tr>
<tr>
<td>80</td>
<td>B</td>
</tr>
<tr>
<td>70</td>
<td>C</td>
</tr>
<tr>
<td>65</td>
<td>D</td>
</tr>
<tr>
<td>0</td>
<td>F</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>≥</th>
<th>Letter</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
<td>A-</td>
</tr>
<tr>
<td>80</td>
<td>B-</td>
</tr>
<tr>
<td>70</td>
<td>C-</td>
</tr>
<tr>
<td>65</td>
<td>D</td>
</tr>
<tr>
<td>0</td>
<td>F</td>
</tr>
</tbody>
</table>
Running the Assignments

A "standard" x86-64 Linux machine, like gl

A passable amount of memory (2GB-4GB)

Modern but not necessarily cutting edge software

Don’t assume a GPU (if you want to write CUDA yourself, talk to me)

If in doubt, ask first
Running the Project

An x86-64 Linux machine

Memory and hardware constraints lifted (somewhat)

If in doubt, ask first
Programming Languages for Assignments

Use the tools you feel comfortable with

Python+numpy, C, C++, Java, Matlab, ...: OK (straight Python may not cut it)

Libraries: Generally OK, as long as you don’t use their implementation of what you need to implement

Math accelerators (blas, numpy, etc.): OK

If in doubt, ask first
Programming Languages for the Project

Use the tools you feel comfortable with

Python+numpy, C, C++, Java, Matlab, ...: OK (straight Python may not cut it)

Libraries: Use what you want

Math accelerators (blas, numpy, etc.): OK
Late Policy

Everyone has a budget of 10 late days
Late Policy

Everyone has a budget of 10 late days.

If you have them left: assignments turned in after the deadline will be graded and recorded, no questions asked.
Late Policy

Everyone has a budget of 10 late days

If you have them left: assignments turned in after the deadline will be graded and recorded, no questions asked.

If you don’t have any left: still turn assignments in. They could count in your favor in borderline cases.
Late Policy

Everyone has a budget of 10 *late days*

Use them as needed throughout the course

They’re meant for personal reasons and emergencies

Do not procrastinate
Late Policy

Everyone has a budget of 10 *late days*

Contact me privately if an extended absence will occur

*You* must know how many you’ve used