## COMP 786 (Fall 2020) Natural Language Processing



THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

**Mohit Bansal** 

### **Class Info/Logistics**

- COMP 786 'Natural Language Processing'
- \*\*3 UNITS\*\*
- Instructor: Mohit Bansal (FB246, <u>http://www.cs.unc.edu/~mbansal/</u>)
- Time: Wed 10.40am-1.10pm ET
- Room: FB008  $\rightarrow$  ZOOM remote
- Office Hours: Wed 1.10-2.10pm ET (appointment, zoom)
- Course Webpage: <a href="https://www.cs.unc.edu/~mbansal/teaching/nlp-comp786-fall20.html">https://www.cs.unc.edu/~mbansal/teaching/nlp-comp786-fall20.html</a>
  - \*\*Course Email\*\*: <u>nlpcomp786unc@gmail.com</u>
- Your email: check/fwd your connectcarolina xyz@email/live.unc.edu email for my welcome message and send me your preferred email id!

## About Me

- Associate Professor, CS, UNC (joined Fall 2016)
- Research Asst. Professor, TTI-Chicago, 2013-2016
- PhD, UC Berkeley, 2008-2013
- Research Interests:
  - Past: Syntactic parsing, coreference resolution, taxonomy induction, world knowledge and commonsense induction
  - Current: Multimodal and embodied semantics (i.e., language with vision and speech, for robotics); human-like language generation and Q&A/dialogue; interpretable and structured deep learning
- Office FB246
- Webpage: <u>http://www.cs.unc.edu/~mbansal/</u>, Email: <u>mbansal@cs.unc.edu</u>
   Lab: <u>https://murgelab.cs.unc.edu/</u>

#### Your Introductions

Please say your:

#### Name

- Department/degree/major
- Research interests (Why NLP? Past ML/AI/NLP/CV experience? Coding experience?)

#### ▶ Fun fact ☺

### About the Course (and its Goals)

- We will cover lots of interesting NLP topics+chapters+papers, brainstorm, & do fun assignments+projects!
- I will start with some basics of traditional NLP
- Then I will cover some specific, latest research topics, both based on traditional models and newer neural models (also some paper readings on certain topics, presented by students)
- Will also discuss connections of NLP with vision and robotics, and several recent deep learning for NLP models
- Some important lectures on ethics/biases in NLP/ML and tips on academic/research quality paper writing and reviewing
- Brainstorm regularly and code + write up fun/novel projects!

#### Prerequisites

Graduate-level class (with some advanced/senior Ugs)

- Machine learning (COMP562 equivalent) and coding experience is definitely expected and required! (please talk to me class if you haven't yet discussed this with me over email)
- Homeworks, projects, paper/chapter summaries, and paper presentations will all require solid ML foundations/ clarity and coding skills (e.g., linear algebra, diff eqns, logistic regression, supervised/unsup learning setups, classifiers, backprop and MLPs/NNs)
- Moreover, some basic NLP background is highly recommended

### Expectations/Grading (tentative)

- Written homework and programming assignments (20%),
- Midterm Project write-up and presentation (20%)
- Final Project write-up and presentation (30%)
- Paper presentations and written summaries (20%)
- Class brainstorming and participation (10%)

#### Homeworks

- Coding based assignments
- On certain topics covered in class, e.g., word embeddings, entailment classification, parsing, Q&A, summarization/ translation
- Code should be written from scratch (acknowledge any borrowed pieces)
- Preferably in tensorflow/python/pytorch

### Chapter/Paper Written Summaries

- 0.5-1 page (per paper) write-up for certain week's subchapters/papers
- Describe the task
- Summarize the methods/models
- Explain the novelty
- Discuss the next steps or potential improvements

#### **Paper Presentation**

- Lead discussion for 1-2 subchapters/papers on a topic some week (may be done in pairs/groups depending on class size)
- Read related chapters/papers and present background to audience
- Present task and ML details of given chapter/papers
- Present demo's of related code, etc.
- Ask interesting questions to initiate brainstorming
- Mention some next steps, future work, extension ideas!

# Project

- Students will pick (early) their favorite topic among latest cutting-edge research topics covered in class
- And will try a novel idea (implementing+extending or original) --I am happy to discuss details!
- Midterm and final report + presentation (and possibly some intermediate updates)
- Preferably individual but we can discuss pairing too
- Use conference style files and aim for high-quality project writeups
- Will have a lecture on research-quality paper writing

# **Class Participation and Brainstorming**

- Audience students expected to take part in lively discussion in every class and after chapter/paper reading!
- Semi-regularly (i.e., after completing several chapters/papers in 2-3 weeks), we will have a brainstorming and 'idea-generation' session!
- Exact details to be announced soon but students expected to submit and discuss novel idea(s) on the whole general topic, e.g., new related task or dataset, new approach to existing task, combinations of tasks/approaches, etc.
- Don't hesitate to propose fancy ideas ③, but try to keep them grounded/feasible and think of how to approach them realistically (in terms of datasets, models, speed, memory, etc.)

#### **Reference Books**

- SLP2: D. Jurafsky & James H. Martin. "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition". Prentice Hall, Second Edition, 2009.
- SLP3: Many draft chapters of the third edition are available online at <u>https://web.stanford.edu/~jurafsky/slp3/</u>
- FSNLP: Chris Manning and Hinrich Schütze, Foundations of Statistical Natural Language Processing, MIT Press. Cambridge, MA: May 1999. <u>http://nlp.stanford.edu/fsnlp/</u>
- ML Background: Andrew Ng's Coursera Machine Learning course <u>https://www.coursera.org/learn/machine-learning</u>
- Stanford NLP + Deep Learning Class: <u>http://web.stanford.edu/class/cs224n/</u>

# Course Syllabus/Topics (tentative)

- Language Modeling
- Part-of-speech Tagging
- Syntactic Parsing: Constituent, Dependency, CCG, others
- Coreference Resolution
- Distributional Semantics: PMI, neural, CCA
- Compositional Semantics: Logical-form, Semantic Parsing, Vector-form, neural (RNNs/ CNNs)
- Question Answering: Factoid-based, Passage-based
- Sentiment Analysis
- Document Summarization
- Machine Translation
- Dialogue Models
- Language and Vision: Image Captioning, Video Captioning, Visual Question Answering
- Language and Robotics: Instructions for Navigation, Manipulation, Skill Learning; Human-Robot Interaction
- Models: Deep+structured, interpretable, adversarial, reward-based (reinforcement learning), etc.
- Ethics and Bias in NLP/ML Models
- How to Write and Review Research Papers

#### Question answering



#### Question answering

Google	What was the U.S. population when Bernie Sanders was born?								۹
	All	News	Images	Videos	Shopping	More -	Search tools		
	About	t 1,620,000	results (0.67	seconds)					
	United States of America / Population (1941)								-
	133.4 million							F	North Pacific Ocean
							Feedba	ck	

#### Question answering

Google	whic	<b>୍</b>						
	All	Maps	Images	News	Shopping	More <b>•</b>	Search tools	

About 2,710,000 results (0.81 seconds)

This major inland sea is bordered by six countries — **Romania** and **Bulgaria** to the west; **Ukraine**, **Russia**, and **Georgia** to the north and east; and **Turkey** to the south. Additionally, it is impacted by another 10 nations through the five major rivers that empty into the Black Sea, the largest of which is the Danube River.



Black Sea Geography - College of Earth, Ocean, and Environment https://www.ceoe.udel.edu/blacksea/geography/index.html University of Delaware -

About this result . Feedback

#### Machine Translation



Yaha maśīna anuvāda kā ēka udāharaņa hai

#### Sentiment Analysis



#### Sentiment Analysis with Python NLTK Text Classification

This is a demonstration of **sentiment analysis** using a NLTK 2.0.4 powered **text classification** process. It can tell you whether it thinks the text you enter below expresses **positive sentiment**, **negative sentiment**, or if it's **neutral**. Using **hierarchical classification**, *neutrality* is determined first, and *sentiment polarity* is determined second, but only if the text is not neutral.

#### Analyze Sentiment

Language

english 🖨

#### Enter text

Analyze

It always amazes me how Universal never cares to create anything remotely clever when it comes to their animations, and so once again they come up with a harmless little story that wants to be cute and funny (which it is sometimes) but is only bound to be quickly forgotten.

Enter up to 50000 characters

#### **Sentiment Analysis Results**

The text is neg.

The final sentiment is determined by looking at the classification probabilities below.

#### Subjectivity

- neutral: 0.3
- polar: 0.7

#### Polarity

- pos: 0.2
- neg: 0.8

#### Natural Language Generation: Summarization





Natural Language Generation: Conversation/Dialogue



#### Multimodal Natural Language Generation: Image Captioning

C 🛈 www.cs.toronto.edu/~nitish/nips2014demo/results/8832804.html

#### Results

#### Tags

- authors
- scones
- luncheon
- breakfast
- seder

#### Nearest Caption in the Training Dataset

a man cuts a cake while children sit around at the table , looking on .

#### **Generated Captions**

- two people at a table with a cake .
- the two people are having a meal that is in a party.
- a man and two children in a blue table with a cake .
- a man sitting at a table with a bunch of cake on it .
- a man and woman sitting at a table with cake at a party .



#### <u>back</u>

#### Multimodal Natural Language Generation: Video Captioning



Ground truth: Two women are shopping in a store. Two girls are shopping.Baseline model: A man is doing a monkey in a store.Multi-task model: A woman is shopping in a store.



Ground truth: Two men are fighting. A group of boys are fighting. Baseline model: A group of men are dancing. Multi-task model: Two men are fighting.



Ground truth: A man is playing a violin. A man is playing the violin on stage.

**Baseline-XE:** A man is playing the drums. **CIDEr-RL:** A man is playing a guitar. **CIDEnt-RL:** A man is playing a violin.



Ground truth: Two men are wrestling. Two guys are wrestling in a competition. Baseline-XE: A group of people are playing a game. CIDEr-RL: A man is playing a wrestling. CIDEnt-RL: Two men are wrestling.



Ground truth: A person is playing a video game. Someone is playing video game. Baseline-XE: A man is riding a motorcycle. CIDEr-RL: A man is talking about a plane. CIDEnt-RL: A person is playing a video game.

#### [Pasunuru and Bansal, 2017]

#### Multimodal NLU: Visual Question Answering



What color are her eyes? What is the mustache made of?



How many slices of pizza are there? Is this a vegetarian pizza?



Is this person expecting company? What is just under the tree?



Does it appear to be rainy? Does this person have 20/20 vision?

Multimodal NLU: Video+Subtitle Question Answering



Question: What is <u>Sheldon</u> holding when he is talking to Howard about the sword? Correct Answer: A computer.



Question: Who is talking to Howard when he is in the kitchen upset? Correct Answer: Raj is talking to Howard.





[Anderson et al., 2017; Chen et al., 2019]

Automatic Speech Recognition



#### Some Exciting NLP Challenges

Non-literal: Idioms, Metaphors



You: *I am under the weather today.* Siri: *Here is the weather today...50 F* 

Non-literal: Idioms, Metaphors



Humor, Sarcasm, Politeness/Rudeness



Prepositional Attachment, Coreference Ambiguities



Prepositional Attachment, Coreference Ambiguities



# Visually Grounded Language

Text-Image Alignment: Most of our daily communication language points to several objects in the visual world



# Visually Grounded Language

Visual Question Answering: Humans asking machines about pictures/videos, e.g., for visually impaired, in remote/ dangerous scenarios, in household service settings





# Embodied Language (Robot Instructions)

Task-based instructions, e.g., navigation, grasping, manipulation, skill learning



# Embodied Language (Robot Instructions)

Task-based instructions, e.g., navigation, grasping, manipulation, skill learning



# Grounded Language Generation/Dialogue

Both for answering human questions, and to ask questions back, and for casual chit-chat



# Grounded Language Generation/Dialogue

Both for answering human questions, and to ask questions back, and for casual chit-chat



#### 10-min break?

#### Language Modeling and Generation

(some slides adapted/borrowed from courses by Dan Klein, Chris Manning, Richard Socher)

A language model is a distribution over sequences of words (sentences)

 $\mathsf{P}(\mathbf{w}) = \mathsf{P}(\mathsf{w}_{1 \dots} \mathsf{w}_{n})$ 

- Purpose is to usually assign high weights to plausible sentences, e.g., in speech recognition or machine translation
- Also used for language generation now (predict next word given previous words), esp. w/ new RNN models

## **Traditional N-gram LMs**

Use chain rule to generate words left-to-right

$$P(w_1 \dots w_n) = \prod_i P(w_i | w_1 \dots w_{i-1})$$

Can't condition on the entire left context

P(??? | Turn to page 134 and look at the picture of the)

N-gram models make a Markov assumption

$$P(w_1 \dots w_n) = \prod_i P(w_i | w_{i-k} \dots w_{i-1})$$

P(please close the door) =

 $P(\text{please}|\text{START})P(\text{close}|\text{please})\dots P(\text{STOP}|door)$ 

## **Traditional N-gram LMs**

- How do we know P(w | history)?
  - Use statistics from data (examples using Google N-Grams)
  - E.g. what is P(door | the)?

198015222 the first 194623024 the same 168504105 the following 158562063 the world 14112454 the door

23135851162 the \*

 $\hat{P}(\text{door}|\text{the}) = \frac{14112454}{23135851162}$ = 0.0006

This is the maximum likelihood estimate

# Sparsity Issue & Parameter Estimation

- New words all the time (antidisestablishmentarianism, kakorrhaphiophobia,, www.xyzabc156.com)....worse for new bigrams and trigrams!
  - Maximum likelihood estimates won't get us very far

$$\hat{P}(w|w_{-1}) = \frac{c(w_{-1}, w)}{\sum_{w'} c(w_{-1}, w')}$$

- Need to smooth these estimates
- General method (procedurally)
  - Take your empirical counts
  - Modify them in various ways to improve estimates
- General method (mathematically)
  - Often can give estimators a formal statistical interpretation
  - ... but not always
  - Approaches that are mathematically obvious aren't always what works

# **Smoothing Techniques**

#### We often want to make estimates from sparse statistics:

- P(w I denied the) 3 allegations 2 reports 1 claims 1 request 7 total 4 total
- Smoothing flattens spiky distributions so they generalize better





Very important all over NLP, but easy to do badly!

# **Smoothing Techniques**

- Classic Solution: add-one or add small priors to numer/denom
- Backing off to smaller n-grams
- Held-out Reweighting: Important to optimize/estimate how models generalize! So use held-out data to estimate the map of old count to new count
- Kneser-Ney Discounting: two successful ideas:
  - Idea 1: observed n-grams occur more in training than they will later
  - Idea 2: Type-based fertility (based on how common the word type is)
- Read Chen and Goodman, 1996 for various details and graphs!

## **RNN Language Models**

- Avoid huge number of n-grams; Memory requirement only scales with #words
- Can condition on all previous history (with forget gates)
- Loss function on identity of predicted word at each time step
- But harder/slower to train and reach optimum (and less interpretable)?



## **Distributional Semantics**

- Words occurring in similar context have similar linguistic behavior (meaning) [Harris, 1954; Firth, 1957]
- Traditional approach: context-counting vectors
  - Count left and right context in window
  - Reweight with PMI or LLR
  - Reduce dimensionality with SVD or NNMF

[Pereira et al., 1993; Lund & Burgess, 1996; Lin, 1998; Lin and Pantel, 2001; Sahlgren, 2006; Pado & Lapata, 2007; Turney and Pantel, 2010; Baroni and Lenci, 2010]

More word representations: hierarchical clustering based on bigram LM LL [Brown et al., 1992]

.011 001 010, 101 110 Apple IBM bought run of in apple pear

0.6

-0.2

0.9

0.3 -0.4

0.5

#### **Unsupervised Embeddings**

Vector space representations learned on unlabeled linear context (i.e., left/right words): distributional semantics (Harris, 1954; Firth, 1957)



## **Distributional Semantics -- NNs**

Newer approach: context-predicting vectors (NNs)

SENNA [Collobert and Weston, 2008; Collobert et al., 2011]: Multi-layer DNN w/ ranking-loss objective; BoW and sentence-level feature layers, followed by std. NN layers. Similar to [Bengio et al., 2003].



# **Distributional Semantics -- NNs**

CBOW, SKIP, word2vec [Mikolov et al., 2013]: Simple, super-fast NN w/ no hidden layer. Continuous BoW model predicts word given context, skipgram model predicts surrounding context words given current word



Other: [Mnih and Hinton, 2007; Turian et al., 2010]

Demos: <u>https://code.google.com/p/word2vec</u>, <u>http://metaoptimize.com/projects/wordreprs/</u>, <u>http://ml.nec-labs.com/senna/</u>

#### **Distributional Semantics**

- Other approaches: spectral methods, e.g., CCA
  - Word-context correlation [Dhillon et al., 2011, 2012]
  - Multilingual correlation [Faruqui and Dyer, 2014; Lu et al., 2015]
- Some later ideas: Train task-tailored embeddings to capture specific types of similarity/semantics, e.g.,
  - Dependency context [Bansal et al., 2014, Levy and Goldberg, 2014]
  - Predicate-argument structures [Hashimoto et al., 2014; Madhyastha et al., 2014]
  - Lexicon evidence (PPDB, WordNet, FrameNet) [Xu et al., 2014; Yu and Dredze, 2014; Faruqui et al., 2014; Wieting et al., 2015]
  - Combining advantages of global matrix factorization and local context window methods [GloVe; Pennington et al., 2014]

- Composing, combining word vectors to representations for longer units: phrases, sentences, paragraphs, …
- Initial approaches: point-wise sum, multiplication [Mitchell and Lapata, 2010; Blacoe and Lapata, 2012]
- Vector-matrix compositionality [Baroni and Zamparelli, 2010; Zanzotto et al., 2010; Grefenstette and Sadrzadeh, 2011; Socher et al., 2011; Yessenalina and Cardie, 2011]
- Linguistic information added via say parses in RvNNs [Socher et al., 2011b, 2012, 2013a, 2013b, 2014; Hermann and Blunsom, 2013]
- Sequential RNNs (with GRU/LSTM gates) (Simple vector averaging w/ updating sometimes competitive)

Feed-forward NNs with back-propagation

Softmax (= logistic regression) is not very powerful

• Softmax only linear decision boundaries



Feed-forward NNs with back-propagation

#### **Neural Nets for the Win!**

• Neural networks can learn much more complex functions and nonlinear decision boundaries!





Feed-forward NNs with back-propagation

A neuron is essentially a binary logistic regression unit

h<sub>w,b</sub>(x)

$$h_{w,b}(x) = f(w^{\mathsf{T}}x + b) \longleftarrow$$

$$f(z) = \frac{1}{1 + e^{-z}}$$

X3.

*b:* We can have an "always on" feature, which gives a class prior, or separate it out, as a bias term



*w*, *b* are the parameters of this neuron i.e., this logistic regression model

Feed-forward NNs with back-propagation

A neural network

= running several logistic regressions at the same time

Before we know it, we have a multilayer neural network....



Feed-forward NNs with back-propagation

**Training with Backpropagation** 

• Let's consider the derivative of a single weight  $W_{ij}$ 

$$\frac{\partial s}{\partial W} = \frac{\partial}{\partial W} U^T a = \frac{\partial}{\partial W} U^T f(z) = \frac{\partial}{\partial W} U^T f(Wx + b)$$

- This only appears inside a<sub>i</sub>
- For example:  $W_{23}$  is only used to compute  $a_2$



Feed-forward NNs with back-propagation 

**Training with Backpropagation**  $\frac{\partial s}{\partial W} = \frac{\partial}{\partial W} U^T a = \frac{\partial}{\partial W} U^T f(z) = \frac{\partial}{\partial W} U^T f(Wx + b)$  $\frac{\partial y}{\partial x} = \frac{\partial y}{\partial u} \frac{\partial u}{\partial x}$ Derivative of weight  $W_{ii}$ :  $z_i = W_{i.}x + b_i = \sum_{i=1}^{3} W_{ij}x_j + b_i$  $\frac{\partial}{\partial W_{ii}} U^T a \rightarrow \frac{\partial}{\partial W_{ii}} U_i a_i$  $a_i = f(z_i)$  $U_{2}$  $U_i \frac{\partial}{\partial W_{ij}} a_i = U_i \frac{\partial a_i}{\partial z_i} \frac{\partial z_i}{\partial W_{ij}}$  $= U_i \frac{\partial f(z_i)}{\partial z_i} \frac{\partial z_i}{\partial W_{\cdot\cdot}}$  $W_{23}$  $= U_i f'(z_i) \frac{\partial z_i}{\partial W_{ii}}$  $= U_i f'(z_i) \frac{\partial W_i \cdot x + b_i}{\partial W \cdot \cdot}$ 

າດ

Feed-forward NNs with back-propagation

**Training with Backpropagation** 

Derivative of single weight  $W_{ij}$ :  $z_i = W_{i,x} + b_i = \sum_{i=1}^{\infty} W_{ij}x_j + b_i$  $U_i \frac{\partial}{\partial W_{ii}} a_i = U_i f'(z_i) \frac{\partial W_i \cdot x + b_i}{\partial W_{ii}}$  $a_i = f(z_i)$  $= U_i f'(z_i) \frac{\partial}{\partial W_{ij}} \sum_{k} W_{ik} x_k$  $= \underbrace{U_i f'(z_i)}_{\delta_i} x_j$  $= \underbrace{\delta_i}_{\mathbf{v}_j} x_j$  $U_2$  $a_1$  $W_{23}$ Local input Local error signal signal where f'(z) = f(z)(1 - f(z)) for logistic f

#### **Recurrent NNs**

- Recurrent NNs (RNNs) are non-tree, sequential versions of recursive RvNNs
- Weights tied together for each time step
- Loss function on identity of predicted word at each time step



### LSTM RNNs

LSTM (Long short term memory) RNNs have gates for forgetting, allowing learning of longer-term connections by avoiding vanishing/ exploding gradients



#### **Character RNNs**

- Can directly process each character as a unit!
- Helps learn prefixes, stems, suffixes (form vs. function, rare/ unseen words, etc.)



#### **RNN** Generations

#### Automatically generate Shakespeare from RNNs!

#### PANDARUS:

Alas, I think he shall be come approached and the day When little srain would be attain'd into being never fed, And who is but a chain and subjects of his death, I should not sleep.

#### Second Senator:

They are away this miseries, produced upon my soul, Breaking and strongly should be buried, when I perish The earth and thoughts of many states.

#### DUKE VINCENTIO: Well, your wit is in the care of side and that.

Second Lord: They would be ruled after this chamber, and my fair nues begun out of the fact, to be conveyed, Whose noble souls I'll have the heart of the wars.

#### Clown: Come, sir, I will make did behold your worship.

VIOLA: I'll drink it.

#### **RNN** Generations

#### Automatically generate Wikipedia-style text from RNNs!

Naturalism and decision for the majority of Arab countries' capitalide was grounded by the Irish language by [[John Clair]], [[An Imperial Japanese Revolt]], associated with Guangzham's sovereignty. His generals were the powerful ruler of the Portugal in the [[Protestant Immineners]], which could be said to be directly in Cantonese Communication, which followed a ceremony and set inspired prison, training. The emperor travelled back to [[Antioch, Perth, October 25/21]] to note, the Kingdom of Costa Rica, unsuccessful fashioned the [[Thrales]], [[Cynth's Dajoard]], known in western [[Scotland]], near Italy to the conquest of India with the conflict. Copyright was the succession of independence in the slop of Syrian influence that was a famous German movement based on a more popular servicious, non-doctrinal and sexual power post. Many governments recognize the military housing of the [[Civil Liberalization and Infantry Resolution 265 National Party in Hungary]], that is sympathetic to be to the [[Punjab Resolution]] (PJS)[http://www.humah.yahoo.com/quardian. cfm/7754800786d17551963s89.htm Official economics Adjoint for the Nazism, Montgomery was swear to advance to the resources for those Socialism's rule, was starting to signing a major tripad of aid exile.

#### **RNN** Generations

Automatically generate source code from RNNs!

```
{ { cite journal | id=Cerling Nonforest Department | format=Newlymeslated | none } }
''www.e-complete''.
'''See also''': [[List of ethical consent processing]]
== See also ==
*[[Iender dome of the ED]]
*[[Anti-autism]]
===[[Religion|Religion]]===
*[[French Writings]]
*[[Maria]]
*[[Revelation]]
*[[Mount Agamul]]
== External links==
* [http://www.biblegateway.nih.gov/entrepre/ Website of the World Festival. The labour
==External links==
* [http://www.romanology.com/ Constitution of the Netherlands and Hispanic Competition
```

### Various Applications of such RNNs

- Language Modeling and Language Generation
- Classification: Sentiment Analysis
- Conditioned Generation: End-to-end MT, Summarization
- Others: Parsing, Captioning, Q&A, Dialogue (some will be covered in future weeks)