First Midterm Exam

601.467/667 Introduction to Human Language Technology

Fall 2021
Johns Hopkins University
Co-ordinator: Philipp Koehn

7 October 2021

Complete all questions.

Use additional paper if needed.

Time: 75 minutes.

Name of student: Yuyang Zhou

Q1. Language Models and Morphology *20 points*

1. (2 points) I have a **unigram** word model trained on standard English. How would you expect the perplexity of the model on the sentence 'a car David rented' compare to the perplexity of the same model on 'David rented a car'? Circle one:

   (a) The perplexity on 'a car David rented' is greater

   (b) The perplexities are equal

   (c) The perplexity on 'David rented a car' is greater

2. (2 points) I have a **bigram** word model trained on standard English. How would you expect the perplexity of the model on the sentence 'a car David rented' compare to the perplexity of the same model on 'David rented a car'? Circle one:

   (a) The perplexity on 'a car David rented' is greater

   (b) The perplexities are equal

   (c) The perplexity on 'David rented a car' is greater

3. (2 points) You have an n-gram distribution which you smooth using add-ϵ, for some $\epsilon > 0$. The entropy of the smoothed distribution is:

   (a) smaller than

   (b) equal to

   (c) greater than

   the entropy of the original distribution (circle one above).

4. (2 points) You are told to build a statistical machine translation system between English and Finnish using only 10 million words of Finnish-English bitext and all the monolingual text data found in Finnish and English Google Books data.

   Which translation direction is likely to achieve better performance on a word-based bleu score (circle one):

   (a) English-to-Finnish

   (b) Finnish-to-English

5. (4 points) Give at least 2 reasons **WHY**, based on what you know about both language modeling and the morphology of Finnish and English:

   1. Based on BLEU scores, Finnish is a morphologically rich language, and it is always harder to translate into such a language than morphologically poor language, such as English.

   2. For the same number of words, English has much fewer word types than Finnish and is more orderly. So training an English language model is easier and can get more accurate information.

6. (2 points) Bound morphemes are (circle one):

   (a) Words or morphemes that keep the same form every time used and are unchangeable, including conjunctions

   (b) Morphemes that cannot stand alone as a word, and must be attached to a free morpheme

   (c) Words that have morphemes that change depending on the grammar and meaning of a sentence, including nouns

7. (3 points) Give the meaning of the word *cooler*:

   (a) when *-er* is an inflectional morpheme feeling moderately cold (adj.)

   (b) when *-er* is a derivational morpheme Something that cools or makes cool (n.)

8. (3 points) give at least 3 distinct allomorphs for the English negative prefix *in-*, also simply describe the orthographic context in which the use of each is typically observed, and give one example word for each:

	Allomorph	Orthographic Context	Example Word
1	[in]	when proceeding an alveolar consonant	intolerate
2	[iŋ]	when proceeding a velar consonant	incongruous
3	[im]	when proceeding a bilabial consonant	impossible

Q2. Syntax *20 points*

(5 points) Explain the difference between *constituency* and *dependency* grammars. Do your best to illustrate each kind of parse of the sentence *She plucked a white flower.*

Difference: While parsing, dependency grammars display only relationships between words and their constitutes. But constituency grammars display the entire sentence structure and relationships.

1. constituency parsing: [S[NP[NN 'She']][VP[VB 'plucked'][NP[DT 'a'][NP[ADJ 'white'][NN 'flower']]]]]]
2. dependency parsing: See above.

(3 points) What is the head of a phrase? What are some of the things it does? Give two examples.

Head of a phrase: The sub-consititute which determines the interal structure and external distribution of the constituent as a whole.
Head often controls the structure of its modifications and has dependents(can be required or optional)
Examples: 1. In sentence: usually the main verb. eg. She **dropped** a book.
2. In noun verb: usually the main verb. eg. the white **flower**

(2 points) In the following example, both *see* and *give* are verbs. Which sentence is ungrammatical? Why?

- Kim planned to give Sandy books.
- Kim planned to see Sandy books.

The second one is ungrammatical. Should be "Kim planned to see Sandy's books."

(5 points) Below is a grammar (a '|' represents multiple options). If derivations start with the S node, list five sentences that could be produced by this grammar. You don't need to show the parse structures, but it doesn't hurt.

S → NP VP
NP → NN
NP → DT NN
NP → DT ADJ NN
VP → VB NP

Dog eats the bone.
A boy picks the green peony.
The generous boy redefined the peony.
A bot picks the cotton-headed bone.
The Dog eats cotton-headed peony.

DT → a | an | the
ADJ → generous | cotton-headed | green
NN → dog | peony | bone | boy
VB → eats | redefines | picks

(5 points) How could you extend this grammar to support plural nouns and verbs, such that agreement is enforced? (i.e., plural nouns can only appear with plural verbs, likewise for singular). You should need at least five more rules. Give at least one new sentence from your new grammar.

Add the following rules into original rules:

S -> NPS VPS
VPS -> VBS NP
VPS -> VBS NPS
NPS -> NNS
NPS -> DTS NNS
NPS -> DTS ADJ NNS
DTS -> the
NNS -> dogs | toys | girls | boys
VBS -> pick | eat | like | see

New sentence:

Dogs like the cotton-headed toys.
The girls see a green peony.

Q3. Semantics *15 points*

(5 points) Explain the difference between *polysemy* and *homonymy*, and illustrate with an example of each.

Difference: For same formd texts, polysemy means many related meanings, homonymy means different meaning.

Examples: 1. Polusemy(hand): Give me a hand please. He is a new hand.
2. Homonymy(prey/pray): On Sunday they pray for you, but on Monday they prey for you.

(5 points) Give an example of two sentences that describe the same *event* but with participants in different *syntactic positions*. Why is this useful information, beyond simply understanding each sentence in isolation?

Examples: Mary visited the book store. The bookstore was visited by Mary.

Because we us Semantic Role Lables to identify synatic position of a sentence, we could determine the word relationships of a sentence, and know it's meaning.

(5 points) You're thinking about getting a new pet (a turtle), but are worried how it would interact with your current pet (a dog). Will they fight? Cuddle? Ignore each other? How could you use the output of a semantic role labeling system to answer this question?

First, we could try to think that animals can communicate using their own languages, which can be generated by their voices. We then use some ways to figure out the Turtle and Dog language by processing their sound or sound wave. Then we could get some sentences they daily "speak". At last, we could use SRL system to idenify synatic position of a sentence, and then know the meaning and get the answer about their interaction and feelings.

Q4. Deep Learning *20 points*

(2 points each) Give 4 examples of Supervised Learning tasks in HLT. Give an example of one input and output that could exist in the trainig data.

Supervised Learning tasks in HLT: 1.Language Identification 2.Speech Recognition 3. Machine Translation 4. Reading Comprehension

Example of input and output: for language identification:
input: (sentence) I eat an apple.
output:(Language) English

(2 points) Why do we use a softmax? (Hint what range is the output of a softmax)

Softmax makes the output a probability distribution, which can be displayed to a user or used as an input to other systems.

(2 points) BERT is a Masked Language Model. What are Masked Language Models and how are they different than n-gram Language Models?

Masked language model is a fill-in-the-blank task, where a model uses the context words surrounding a mask token to predict what the masked word should be. But n-gram model predicts the probability of a given n-gram with any sequence of words in the language.

(4 points) Why have Neural Networks taken over in HLT in the last 10 years? Please give at least 2 reasons.

1. Data availability increased significantly(such as storage capacity).
2. The increasing of performance of GPUs.

(4 points) How are words input into a Neural Network? What sort of problems could this cause and how do we solve them?

Words input: One-Hot-Vector
Problems: 1. The representation size grows with the corpus, so computationally too expensive.
2. Each vector is equidistant from every vector in a one-hot encoding. Sometimes we need distributed representation to capture more information.
Solve: For problem 1, we could lower thr demension by some ways like PCA.
For question 2, we can aviod using one-hot encoding whe we need complex information about an word.

Q5. Information Retrieval *10 points*

You've built a new Information Retrieval system and need to evaluate whether it is good according to the Mean Average Precision (MAP) metric. Suppose you index 1000 documents d_1, d_2, d_3, . . ., d_{1000}, and then try searching with two queries q_1 and q_2, each of which retrieving a ranked list of documents, as follows:

1. Query: q_1

   Ranked list of documents retrieved by your system, in order: d_3, d_4, d_2, d_5
   (i.e. Document d_3 is deemed best by system, follow by d_4, etc. Documents not listed here are deemed irrelevant by the system)

   Answer key: d_2, d_4
   (i.e. These are relevant documents for q_1, determined by a manual annotation)

2. Query: q_2

   Ranked list of documents retrieved by your system, in order: d_3, d_1, d_5, d_9

   Answer key: d_3, d_5, d_6, d_7

Please compute the following quantities. For simplicity, leave numbers in fractional form.

- (1 point) Precision for q_1: `1/2`
- (1 point) Recall for q_1: `1`
- (1 point) Precision for q_2: `1/2`
- (1 point) Recall for q_2: `1/2`
- (2 points) Average Precision for q_1: `7/12`
- (2 points) Average Precision for q_2: `5/6`
- (2 points) MAP for the two queries: `17/24`

Q6. Information Extraction *25 points*

Consider the following text:

> lululemon athletica inc. (LULU) is reporting for the quarter ending July 31, 2021. The textile company's consensus earnings per share forecast from the 11 analysts that follow the stock is $1.21. This value represents a 63.51% increase compared to the same quarter last year. In the past year LULU has beat the expectations every quarter. The highest one was in the 2nd calendar quarter where they beat the consensus by 27.47%. Zacks Investment Research reports that the 2022 Price to Earnings ratio for LULU is 55.21 vs. an industry ratio of 20.10, implying that **they** will have a higher earnings growth than their competitors in the same industry.
>
> Oracle (ORCL) reported quarterly results late Monday that slightly missed revenue estimates and soundly beat on earnings. Oracle stock fell. The database software company reported adjusted earnings of $1.03 a share on revenue of $9.73 billion. Analysts expected Oracle to report earnings of 97 cents on revenue of $9.75 billion, according to FactSet. The results were for its fiscal first quarter ended Aug. 31. Revenue climbed 4% from the year-ago period. Oracle stock fell 3%, near 86.10, during after-hours trading on the stock market today.
>
> Apple today announced financial results for its fiscal 2021 third quarter ended June 26, 2021. The Company posted a June quarter record revenue of $81.4 billion, up 36 percent year over year, and quarterly earnings per diluted share of $1.30.

You are tasked to extract earning statistics from text like this and store it in a database table that contains

- company name
- ticker symbol
- earnings per share
- percentage change in earnings

See the next page.

Entity linking Mark in the text where companies referred to and indicate which mentions are referring to the same company.

Name one feature that would a co-reference resolution method conclude that the pronoun *they* (bold, at the end of the first paragraph) refers to *lululemon athletica inc.*. Also name one feature that would interfere with this conclusion.

References are marked with yellow in the text.
Feature 1: Ψ (they, lululemon athletica inc.) > Ψ(they, industry)
Feature 2: Ψ(they, industry) > Ψ(they, lululemon athletica inc.)

Surface extraction rule Write a surface pattern rule that allows the extraction of the relationship between company name and ticker symbol.

Rule: [company name]([ticker symbol])

Syntactic extraction rule Write a syntactic pattern rule that allows the extraction of values for the "earnings per share" concept.

[company name] <- SUBJ - has - PP-SHARE -> [earnings per share]

Extra Space

Database Table

Company Name	Ticker symbol	Earnings per share	Percentage change in earnings
Lululemon athletica	LULU	$1.21	63.51%
Oracle	ORCL	$1.03	4%
Apple	Apple	$1.30	36%