Introduction to Word Senses & Semantics

Yu Meng
University of Virginia
yumeng5@virginia.edu

Sep 09, 2024

Reminders

- Assignment 1 is due this Wednesday (09/11) 11:59pm!
- Assignment 2 will be released shortly
- Project proposal is due next Friday (guideline will be released soon)

UNIVERSITY of VIRGINIA

Join at
slido.com
#3387 013

Overview of Course Contents

- Week 1: Logistics & Overview
- Week 2: N-gram Language Models
- Week 3: Word Senses, Semantics & Classic Word Representations
- Week 4: Word Embeddings
- Week 5: Sequence Modeling and Transformers
- Week 6-7: Language Modeling with Transformers (Pretraining + Fine-tuning)
- Week 8: Large Language Models (LLMs) & In-context Learning
- Week 9-10: Knowledge in LLMs and Retrieval-Augmented Generation (RAG)
- Week 11: LLM Alignment
- Week 12: Language Agents
- Week 13: Recap + Future of NLP
- Week 15 (after Thanksgiving): Project Presentations

(Recap) Language Modeling

- Language modeling is the core problem in NLP
- Every NLP task can be formulated as language modeling
- (Autoregressive) language models can be used to generate texts
- Language model distributions are estimated (trained) on a training corpus

UNIVERSITY of VIRGINIA

Join at
slido.com
#3387 013

(Recap) N-gram Language Models

- N-gram language models simplifies the (general) language modeling assumption: the probability of a word is only dependent on the previous N−1 words
- Lower-order N-grams (small N) capture less context information/word correlations
- Higher-order N-grams (bigger N) suffer from more sparsity and huge parameter space
- Smoothing techniques can be used to address sparsity in N-gram language models
  - Add-one smoothing
  - Add-*k* smoothing
  - Language model interpolation
  - Backoff

(Recap) Language Model Evaluation

- Training/validation/test split required before training & evaluating language models
- Perplexity measures how “confused” the language model is about the next word
- Lower perplexity on the test set = better language model
- Perplexity is the commonly used intrinsic evaluation metric for language modeling
- Perplexity is practically computed in the log scale

(Recap) How to Evaluate Language Models?

- What language models should be considered “good”?
  - A perfect language model should be able to correctly predict every word in a corpus
  - We hope the language model can assign a high probability to the next word
  - Better language model = “less surprised” by the next word
- Just use the next word probability assigned by a language model as the metric!
- Does the choice of the evaluation corpus matter?

UNIVERSITY of VIRGINIA

Join at
slido.com
#3387 013

(Recap) Training/Validation/Test Corpus

- **Training corpus/set**: The text data we train our models on
- Does it make sense to evaluate language model probability on the training corpus?
- If we evaluate on the training corpus, we will get misleadingly high probabilities for next word prediction -> train-test data leakage
- **Test corpus/set**: A held-out set of data without overlapping with the training set
- We should always evaluate the model performance using the test corpus which measures the model's generalization ability to unseen data!
- Test sets should **NOT** be used to evaluate language models many times for tuning hyperparameters/design choices -> indirectly learn from test set characteristics
- **Validation/development corpus/set (optional)**: Tuning hyperparameters & making design choices before evaluating on the test set

(Recap) Training/Validation/Test Split

- If we have a fixed amount of data, how should we split into train/valid/test sets?
- We want the training set to be as large as possible
- But the validation/test sets should be also reasonably large to yield reliable evaluation results
- The test set should reflect the data/task we aim to apply language models to

(Recap) Perplexity

- Perplexity (abbreviation: PPL) is an **intrinsic** evaluation metric for language models
- PPL = the per-word inverse probability on a test sequence $\boldsymbol{x}_{\text{test}} = [x_1, x_2, \ldots, x_n]$

$$\mathrm{PPL}(\boldsymbol{x}_{\text{test}}) = \sqrt[n]{\prod_{i=1}^{n} \frac{1}{p(x_i | x_{i-N+1}, \ldots, x_{i-1})}}$$

- A lower PPL = a better language model (less surprised/confused by the next word)

$$\mathrm{PPL}(\boldsymbol{x}_{\text{test}}) = \sqrt[n]{\prod_{i=1}^{n} \frac{1}{p(x_i)}}$$

Unigram

$$\mathrm{PPL}(\boldsymbol{x}_{\text{test}}) = \sqrt[n]{\prod_{i=1}^{n} \frac{1}{p(x_i | x_{i-1})}}$$

Bigram

$$\mathrm{PPL}(\boldsymbol{x}_{\text{test}}) = \sqrt[n]{\prod_{i=1}^{n} \frac{1}{p(x_i | x_{i-2}, x_{i-1})}}$$

Trigram

Perplexity can be used to evaluate general language models (e.g., large language models) too

(Recap) Perplexity: Log-Scale Computation

- Computation of PPL in the raw probability scale can cause numerical instability

$$\mathrm{PPL}(\boldsymbol{x}_{\text{test}}) = \sqrt[n]{\boxed{\prod_{i=1}^{n} \frac{1}{p(x_i | x_{i-N+1}, \dots, x_{i-1})}}}$$

Multiplication of many small probability values!

Example: (1/10) ^ 100 = 10^-100 -> risks of underflow (round to 0)

- PPL is usually computed in the log-scale in practice

$$\mathrm{PPL}(\boldsymbol{x}_{\text{test}}) = \exp\left(\log\left(\sqrt[n]{\prod_{i=1}^{n} \frac{1}{p(x_i | x_{i-N+1}, \dots, x_{i-1})}}\right)\right) = \exp\left(-\frac{1}{n}\sum_{i=1}^{n} \boxed{\log p(x_i | x_{i-N+1}, \dots, x_{i-1})}\right)$$

Log probabilities are numerically stable

Example: log(1/10) = -2.3

Intrinsic vs. Extrinsic Evaluation

- **Intrinsic metrics** (e.g., perplexity) directly measure the quality of language modeling per se, independent of any application
- **Extrinsic metrics** (e.g., accuracy) measure the language model's performance for specific tasks/applications (e.g., classification, translation)
- Intrinsic evaluations are good for development to iterate quickly and understand specific properties of the model
- Extrinsic evaluations are essential to validate that the model improves the performance of an application in a real-world scenario
- Both intrinsic and extrinsic evaluations are commonly used to evaluation language models (they may not be always positively correlated!)

UNIVERSITY of VIRGINIA

Join at slido.com #3387 013

Extrinsic Evaluations for SOTA Language Models

Math reasoning, question answering, general knowledge understanding…

🤗 Open LLM Leaderboard

Model	BBH	MATH Lvl 5	GPQA	MUSR	MMLU-PRO
MaziyarPanahi/calme-2.1-rys-78b	59.47	36.4	19.24	19	49.38
MaziyarPanahi/calme-2.2-rys-78b	59.27	37.92	20.92	16.83	48.73
MaziyarPanahi/calme-2.1-qwen2-72b	57.33	36.03	17.45	20.15	49.05
MaziyarPanahi/calme-2.2-qwen2-72b	56.8	41.16	16.55	16.52	49.27
Qwen/Qwen2-72B-Instruct	57.48	35.12	16.33	17.17	48.92
alpindale/magnum-72b-v1	57.65	35.27	18.79	15.62	49.64
meta-llama/Meta-Llama-3.1-70B-Instruct	55.93	28.02	14.21	17.69	47.88
abacusai/Smaug-Qwen2-72B-Instruct	56.27	35.35	14.88	15.18	46.56
MaziyarPanahi/calme-2.2-llama3-70b	48.57	22.96	12.19	15.3	46.74
NousResearch/Hermes-3-Llama-3.1-70B	53.77	13.75	14.88	23.43	41.41
tenyx/Llama3-TenyxChat-70B	49.62	22.66	6.82	12.52	46.78

Figure source: https://huggingface.co/spaces/open-llm-leaderboard/open_llm_leaderboard

University of Virginia

Join at
slido.com
#3387 013

Agenda

- Introduction to Word Senses & Semantics
- Classic Word Representations
- Vector Space Model Basics

Why Care About Word Semantics?

- Understanding word meanings helps us build better language models!
- Recall the example from N-gram lectures:

$$
\begin{aligned}
&[\text{BOS}]\ \text{The cat is on the mat}\ [\text{EOS}]\\
&[\text{BOS}]\ \text{I have a cat and a mat}\ [\text{EOS}]\\
&[\text{BOS}]\ \text{I like the cat}\ [\text{EOS}]
\end{aligned}
$$

$$p(\text{“cat”}\,|\,\text{“the”}) = \frac{2}{3}, \quad p(\text{“mat”}\,|\,\text{“the”}) = \frac{1}{3},$$

- Sparsity: many valid bigram counts are zero – count-based measures do not account for word semantics!
- If we know “cat” is semantically similar to “dog”, then $p(\text{“dog”}\,|\,\text{“the”}) \approx p(\text{“cat”}\,|\,\text{“the”})$

What Types of Word Semantics Exist in NLP?

- **Synonyms**: words with similar meanings
  - “happy” & “joyful”
- **Antonyms**: words with opposite meanings
  - “hot” & “cold”
- **Hyponyms** & **hypernyms**: one word is a more specific instance of another
  - “rose” is a hyponym of “flower”
  - “flower” is a hypernym of “rose”
- **Polysemy**: A single word having multiple related meanings
  - “mouse” can mean small rodents or the device that controls a cursor
- The study of these aspects of word meanings is called **lexical semantics** in linguistics

Lemmas

- **Lemma**: the base or canonical form of a word, from which other forms can be derived
  - “run” “runs” “ran” and “running” all share the lemma “run”
  - “better” and “best” share the lemma “good”
- **Lemmatization**: reducing words to their lemma
  - Allows models to recognize that different forms of a word carry the same meaning
  - An important pre-processing step in early NLP models
  - Contemporary LLMs (sort of) perform lemmatization through tokenization (later lectures!)

Synonyms

- Word that have the same meaning in some or all contexts
- Two words are synonyms if they can be substituted for each other
- Perfect synonym is very rare!
  - Typically, words are slightly different in notions of politeness, connotation, genre/style…
  - “Child” vs. “kid”: “child” is often more formal/neutral; “kid” is more informal/casual
  - “Slim” vs. “skinny”: “slim” is often more positive in connotation than “skinny”
  - “Big” vs. “Large”: “big sister” is a common phrase but “large sister” is not

UNIVERSITY of VIRGINIA

Join at
slido.com
#3387 013

Antonyms

- Words that have opposite meanings
- Gradable antonyms: exist on the ends of a spectrum or scale
  - "Hot" vs. "cold"
  - "Tall" vs. "short"
- Complementary antonyms: the presence of one directly excludes the other
  - "Alive" vs. "dead"
  - "True" vs. "false"
- Relational antonyms: express a relationship between two dependent entities
  - "Teacher" vs. "student"
  - "Buyer" vs. "seller"

UNIVERSITY of VIRGINIA

Join at
slido.com
#3387 013

Hyponyms & Hypernyms

- Describe hierarchical relationships between words based on specificity and generality
- **Hypernym** is a word that is more general/broader in meaning and can encompass a variety of more specific words
- **Hyponym** is a word that is more specific in meaning and falls under a broader category
- “Vehicle” is a hypernym for “car” “bicycle” “airplane” “boat” etc.
- “Car” “bicycle” “airplane” “boat” are hyponyms of “vehicle”
- **Hypernym/hyponym** relationship is usually transitive
  - A is a hypernym of B; B is a hypernym of C => A is a hypernym of C

Polysemy & Senses

- **Polysemy**: a single word has multiple related meanings
  - "**Light**": "This bag is **light**" / "Turn on the **light**" / "She made a **light** comment"
- **Sense**: a particular meaning or interpretation of a word in a given context
- Word relations (e.g., synonyms, antonyms, hypernyms/hyponyms) are defined between word senses!
- **Word sense disambiguation (WSD):** determine which sense of a word is being used in a specific context
  - She went to the **bank** to deposit money
  - She lives by the river **bank**
- WSD can be challenging especially when the context is short/insufficient
  - Is the query "mouse info" looking for a pet or a tool?

Word Sense Disambiguation

WSD can be an interesting/challenging test case even for the latest LLMs

Image generated by GPT4V under the user prompt: *"generate an image of a baseball player caring for his bat in the cave where he lives with all the other bats"*

Figure source: https://lm-class.org/lectures/04%20-%20word%20embeddings.pdf

Word Similarity

- Most words may not have many perfect synonyms, but usually have lots of similar words
  - “cat” is not a synonym of “dog”, but they are similar in meaning

vanish	disappear	9.8
belief	impression	5.95
muscle	bone	3.65
modest	flexible	0.98
hole	agreement	0.3

Word similarity (on a scale from 0 to 10) manually annotated by humans

- We’ll introduce word embeddings to automatically learn word similarity next week!

Figure source: https://web.stanford.edu/~jurafsky/slp3/6.pdf

Word Relatedness & Semantic Field

- **Word relatedness**: the meaning of words can be related in ways other than similarity
  - Functional relationship: “doctor” and “hospital” – doctors work in hospitals
  - Thematic relationship: “bread” and “butter” – often used together in the context of food
  - Conceptual relationship: “teacher” and “chalkboard” – both part of the educational context
- **Semantic field**: a set of words which cover a particular semantic domain and bear structured relations with each other
  - Semantic field of “houses”: door, roof, kitchen, family, bed…
  - Semantic field of “restaurants”: waiter, menu, plate, food, chef…
  - Semantic field of “hospitals”: surgeon, nurse, anesthetic, scalpel…

Connotation

- Subjective/cultural/emotional associations that words carry beyond their literal meanings
  - Youthful (positive) vs. childish (negative)
  - Confident (positive) vs. arrogant (negative)
  - Economical (positive) vs. cheap (negative)
- Connotation can be described via three dimensions:
  - Valence: the pleasantness of the stimulus
  - Arousal: the intensity of emotion provoked by the stimulus
  - Dominance: the degree of control exerted by the stimulus

UNIVERSITY of VIRGINIA

Join at slido.com #3387 013

Connotation

- Valence: the pleasantness of the stimulus
  - High: “happy” / “satisfied”; low: “unhappy” / “annoyed”
- Arousal: the intensity of emotion provoked by the stimulus
  - High: “excited”; low: “calm”
- Dominance: the degree of control exerted by the stimulus
  - High: “controlling”; low: “influenced”

	Valence	Arousal	Dominance
courageous	8.05	5.5	7.38
music	7.67	5.57	6.5
heartbreak	2.45	5.65	3.58
cub	6.71	3.95	4.24

Earliest work on representing words with multi-dimensional vectors!

Figure source: https://web.stanford.edu/~jurafsky/slp3/6.pdf

Join at
slido.com
#3387 013

Agenda

- Introduction to Word Senses & Semantics
- Classic Word Representations
- Vector Space Model Basics

WordNet

- Word semantics is complex (multiple senses, various relations)!
- How did people represent word senses and relations in early NLP developments?
- WordNet: A manually curated large lexical database
- Three separate databases: one each for nouns, verbs and adjectives/adverbs
- Each database contains a set of lemmas, each one annotated with a set of senses
- Synset (synonym set): The set of near-synonyms for a sense
- Word relations (hypernym, hyponym, antonym) defined between synsets

WordNet: https://wordnet.princeton.edu/

WordNet Relations

Relation	Also Called	Definition	Example
Hypernym	Superordinate	From concepts to superordinates	$\mathit{breakfast}^1 \rightarrow \mathit{meal}^1$
Hyponym	Subordinate	From concepts to subtypes	$\mathit{meal}^1 \rightarrow \mathit{lunch}^1$
Instance Hypernym	Instance	From instances to their concepts	$\mathit{Austen}^1 \rightarrow \mathit{author}^1$
Instance Hyponym	Has-Instance	From concepts to their instances	$\mathit{composer}^1 \rightarrow \mathit{Bach}^1$
Part Meronym	Has-Part	From wholes to parts	$\mathit{table}^2 \rightarrow \mathit{leg}^3$
Part Holonym	Part-Of	From parts to wholes	$\mathit{course}^7 \rightarrow \mathit{meal}^1$
Antonym		Semantic opposition between lemmas	$\mathit{leader}^1 \iff \mathit{follower}^1$
Derivation		Lemmas w/same morphological root	$\mathit{destruction}^1 \iff \mathit{destroy}^1$

Noun relations

Relation	Definition	Example
Hypernym	From events to superordinate events	$\mathit{fly}^9 \rightarrow \mathit{travel}^5$
Troponym	From events to subordinate event	$\mathit{walk}^1 \rightarrow \mathit{stroll}^1$
Entails	From verbs (events) to the verbs (events) they entail	$\mathit{snore}^1 \rightarrow \mathit{sleep}^1$
Antonym	Semantic opposition between lemmas	$\mathit{increase}^1 \iff \mathit{decrease}^1$

Verb relations

Figure source: https://web.stanford.edu/~jurafsky/slp3/G.pdf

Join at
slido.com
#3387 013

WordNet as a Graph

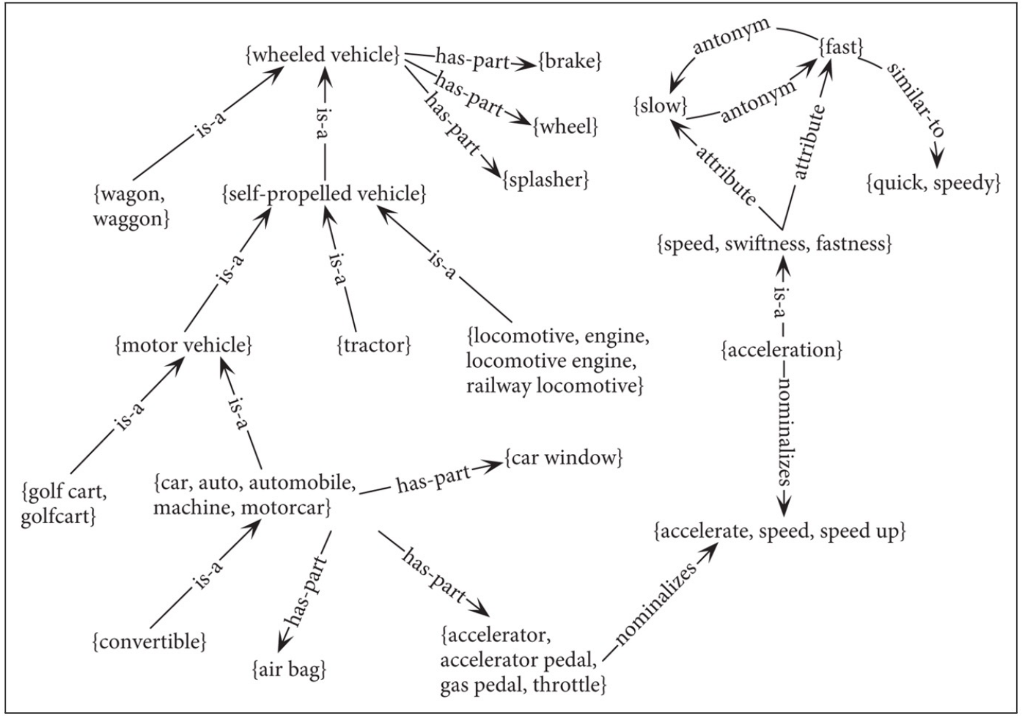

Figure source: https://academic.oup.com/edited-volume/42643/chapter/358151233

UNIVERSITY of VIRGINIA

Join at
slido.com
#3387 013

WordNet Demo

Category	Unique Strings
Noun	117798
Verb	11529
Adjective	22479
Adverb	4481

Figure source: https://lm-class.org/lectures/04%20-%20word%20embeddings.pdf

Word to search for: light [Search WordNet]

Display Options: (Select option to change) [Change]

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Noun

- S: (n) **light**, visible light, visible radiation ((physics) electromagnetic radiation that can produce a visual sensation) *"the light was filtered through a soft glass window"*
  - *direct hyponym* / *full hyponym*
  - *domain category*
  - *direct hypernym* / *inherited hypernym* / *sister term*
  - *part holonym*
  - *derivationally related form*
- S: (n) **light**, light source (any device serving as a source of illumination) *"he stopped the car and turned off the lights"*
- S: (n) **light** (a particular perspective or aspect of a situation) *"although he saw it in a different light, he still did not understand"*
- S: (n) luminosity, brightness, brightness level, luminance, luminousness, **light** (the quality of being luminous; emitting or reflecting light) *"its luminosity is measured relative to that of our sun"*
- S: (n) **light** (an illuminated area) *"he stepped into the light"*
  - *direct hypernym* / *inherited hypernym* / *sister term*
  - *derivationally related form*
- S: (n) **light**, illumination (a condition of spiritual awareness; divine illumination) *"follow God's light"*
- S: (n) **light**, lightness (the visual effect of illumination on objects or scenes as created in pictures) *"he could paint the lightest light and the darkest dark"*
- S: (n) **light** (a person regarded very fondly) *"the light of my life"*
- S: (n) **light**, lighting (having abundant light or illumination) *"they played as long as it was light"; "as long as the lighting was good"*
- S: (n) **light** (mental understanding as an enlightening experience) *"he finally saw the light"; "can you shed light on this problem?"*
- S: (n) sparkle, twinkle, spark, **light** (merriment expressed by a brightness or gleam or animation of countenance) *"he had a sparkle in his eye"; "there's a perpetual twinkle in his eyes"*
- S: (n) **light** (public awareness) *"it brought the scandal to light"*
- S: (n) Inner Light, **Light**, Light Within, Christ Within (a divine presence

WordNet web browser: http://wordnetweb.princeton.edu/perl/webwn

WordNet for Word Sense Disambiguation

- All words WSD task: map all input words (nouns/verbs/adjectives/adverbs) to WordNet senses
- Strong baseline: map to the first sense in WordNet (most frequent)
- Modern approaches: sequence modeling architectures (later lectures!)

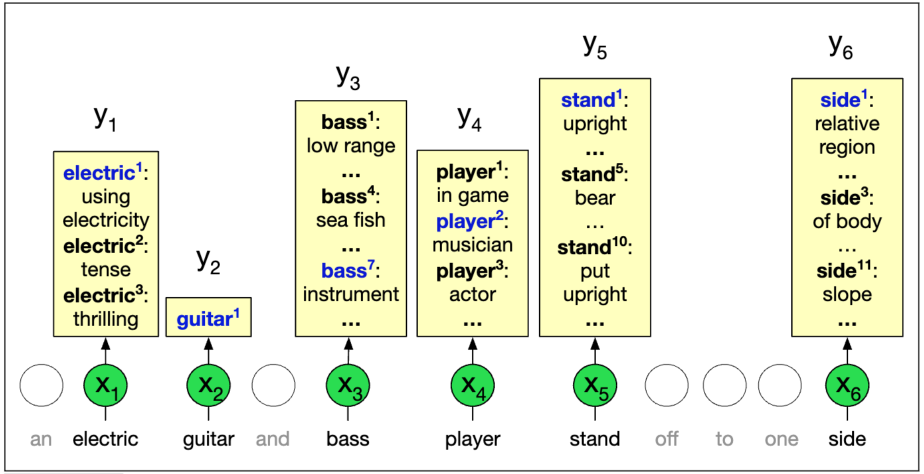

Figure source: https://web.stanford.edu/~jurafsky/slp3/G.pdf

Thank You!

Yu Meng
University of Virginia
yumeng5@virginia.edu