

Exam



925156

Ge Yang

(Q1. a) "Sparsity" means zero probability for any n-gram that has not seen before. The sequence calculation cannot continue if there is zero probability.

(Q1 b) "BLEU" and "ROUGHT" is similar in they both compute overlap of words/n-gram in a sentence. "BLEU" is precision based. but "ROUGHT" is recall based.

(Q1 c) Contextual Representation is representation based on context. It changes for different context and Word2Vec Embedding does not change based on context. It captures multiple senses of a word.

92515G in a limited way

Q1.(d) "Copy" is useful¹ in translation since it allows the algorithm to generate words from source to better summarise the article.

For example, french words sometimes could mean better when translated to English.

But majority English translation should be English so "Copy" is not that useful and could produce some translation not being understandable.

In summary, Copy should not be used but in a very limited situation.

925156

(a)

Q2. Regular Languages are closed means if we take regular language L_1, L_2 to merge them, concatenate them, negate them, intersect them, the results are still formal language. This is important that it allows versatile properties of languages to be used together.

Q2(b). FST does not only accept and score strings which can be done by FSA. FST also intends to translate strings into another language for the purpose of grammar correction & structure parsing. For example, it helps to resolve deflectional morphology. ↓

FST is a kind of Finite state automata, Finite state automata includes FST & FSA.

925156

Q3. (a) α is the prior of topic-word distribution while β is the prior of document-topic distribution.

β should be generally small since we want each topic to focus on specific themes.

α is relatively larger since we expect multiple topics within a document.

(b) Topic model is difficult in:

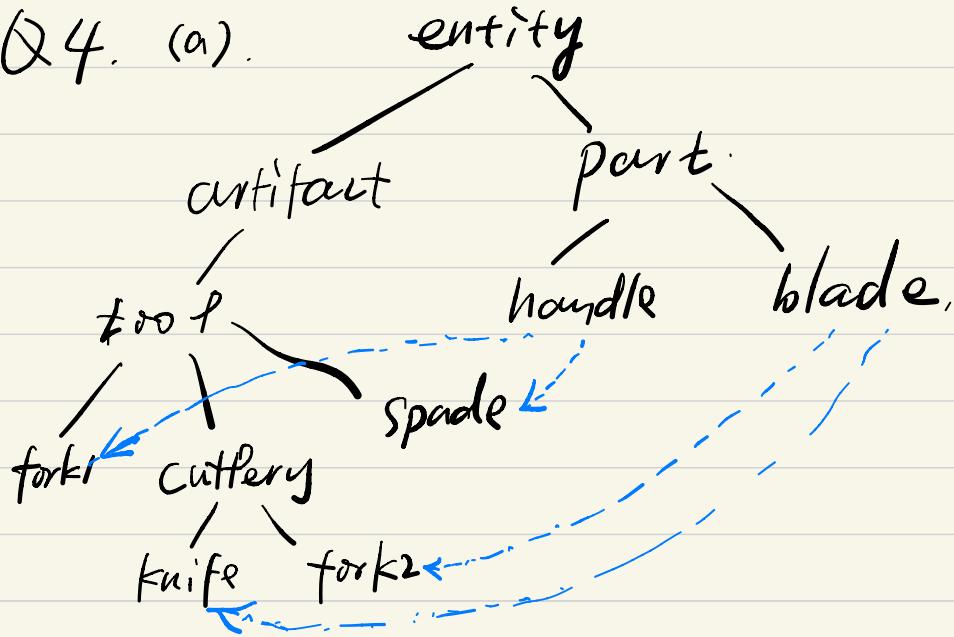
no labels for unsupervised learning.

Perplexity is not comparable for different corpora, different tokenization/preprocessing methods. It also does not correlate with human perception of topic quality.

One method for evaluation is intrinsic evaluation using perplexity. Other methods include Topic Coherence with word intrusion.

925156

Q4. (a).



Q4 (b). LSA is count based method which truncates the matrices in SVD so that lower dimension is used. to tackle sparsity problem with document as context; skip-gram W2V is a neural model which predicts words in local context surrounding given words. It tries to make embedding of neighbouring words similar to that of target words. Skip-gram uses dot product for comparison while LSA uses other similarity. LSA uses document as context while W2V uses whole context as input.

Q5 (a).

Hugh / B-PER Jackman / I-PER is / 0
an / 0 actor / 0 born / 0 in / 0
Sydney / B-LOC NSW / B-LOC Australia / B-LOC

(b) Positive Instances:

$$C_5^2 = \frac{5 \times 4}{2 \times 1} = 10$$

Negative Instances:

$$C_{11}^2 - C_5^2 = 55 - 10 = 45$$

(c) Such that abbreviation "NSW" of "New south whales" in the sentence, they cannot match. Hence, a substitution should be carried out to replace "NSW" by the full Entity Name.

9/25/156

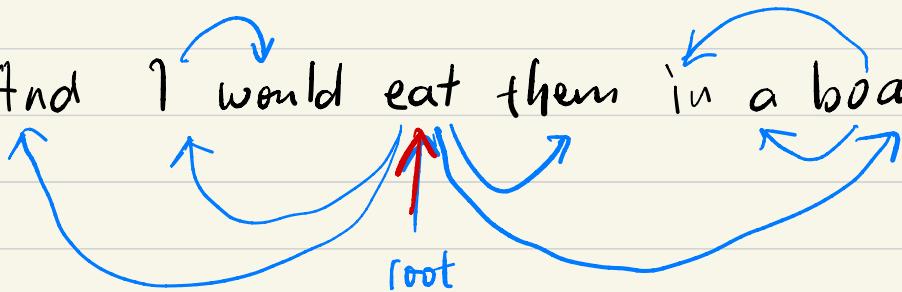
Q6 (a). One word is a head and one word is dependent. There is grammatical relations between these two words.

e.g.: I prefer a meal

meal is dependent & prefer is the head.

Q6
(b).

And I would eat them in a boat



Q6 (c).

Stack WL

Actions Relation

[root] [And I would eat them] Shift

[root. And] [I, would ...] Shift

[root. And I] [would ...] Shift.

[root. And I ~~would~~] [eat them ...] Right Arc \rightarrow world

[root And ~~I~~ eat] [them in ...] Left Arc eat \rightarrow I

~~[root And eat]~~ [them in ...] Left Arc eat \rightarrow Ad

925/56

Q7

$$(a) P(\text{natural} | \langle s \rangle) = \frac{3+1}{3+9} = \frac{1}{3}$$

$$P(\text{language} | \text{natural}) = \frac{3+1}{3+9} = \frac{1}{3}$$

$$P(\text{processing} | \text{language}) = \frac{1+1}{3+9} = \frac{1}{6}$$

$$P(s | s) | \text{processing} = \frac{1+1}{1+9} = \frac{1}{5}$$

$$P(\text{understanding} | \text{language}) = \frac{1+1}{3+9} = \frac{1}{6}$$

$$P(\text{applications} | \text{language}) = \frac{1+1}{3+9} = \frac{1}{6}$$

$$P(in | \text{applications}) = \frac{1+1}{3+9} = \frac{1}{5}$$

$$P(the | in) = \frac{1+1}{1+9} = \frac{1}{5}$$

$$P(wild | the) = \frac{1+1}{1+9} = \frac{1}{5}$$

$$P(</s> | wild) = \frac{1+1}{1+9} = \frac{1}{5}$$

Q7 (b).

$$\begin{aligned} & P(\text{sent} = \text{language understanding} \mid \text{applied}) \\ &= P(\text{languages} \mid \text{ss}) P(\text{understanding} \mid \text{language}) * \\ &\quad P(\text{applications} \mid \text{understanding}) * \\ &\quad P(\text{ss} \mid \text{applications}) \\ &= \frac{0+1}{3+9} \times \frac{1}{6} \times \frac{0+1}{1+9} \times \frac{0+1}{1+9} \\ &= \frac{1}{12} \times \frac{1}{6} \times \frac{1}{10} \times \frac{1}{10} = \frac{1}{7200} \end{aligned}$$

Q7 (c) Kneser-Ney redistributed probability

mass based on continuation number.

Kneser-Ney give probability to other n-gram unseen.

For example :

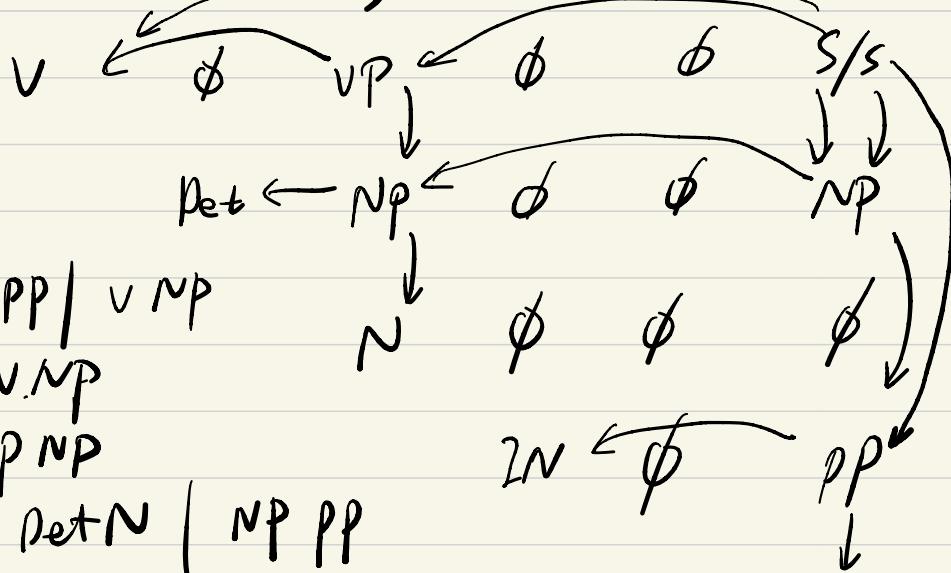
Q8.

a) Find the boy with an Eye.

First meaning is find the boy who has only one eye. Second meaning is to find the boy using one eye.

c)

Find the boy with an eye.



Pt \rightarrow the/an

N \rightarrow boy/eye

V \rightarrow find

P \rightarrow with

Pt \leftarrow NP

NP

NP