Natural Language Processing & Word Embeddings

10/10 分 (100%)

测验, 10 个问题

✔ 恭喜!您通过了!	下一项
1/1分	
1。 Suppose you learn a word embedding for a vocabula 10000 words. Then the embedding vectors should be dimensional, so as to capture the full range of variati meaning in those words.	10000
True	
False	
正确	
1/1分	
2.	
What is t-SNE?	
A linear transformation that allows us to solv analogies on word vectors	re
A non-linear dimensionality reduction techni	que
正确	
A supervised learning algorithm for learning embeddings	word
An open-source sequence modeling library	

Natural Language Processing & Word Embeddings

10/10 分 (100%)

测验, 10 个问题

3。

Suppose you download a pre-trained word embedding which has been trained on a huge corpus of text. You then use this word embedding to train an RNN for a language task of recognizing if someone is happy from a short snippet of text, using a small training set.

x (input text)	y (happy?)
I'm feeling wonderful today!	1
I'm bummed my cat is ill.	0
Really enjoying this!	1

Then even if the word "ecstatic" does not appear in your small training set, your RNN might reasonably be expected to recognize "I'm ecstatic" as deserving a label y=1 .



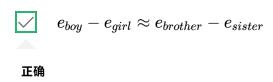


1/1分

False

4

Which of these equations do you think should hold for a good word embedding? (Check all that apply)



	e_{boy} -	– $e_{girl}pprox$	e_{sister}	$-e_{brother}$
--	-------------	-------------------	--------------	----------------

Natural Language Processing & Word Embeddings 未选择的是正确的

10/10 分 (100%)

测验, 10 个问题

✓	$e_{boy}-e_{brother}$	\approx	e_{girl} –	e_{sister}
正确				

未选择的是正确的



1/1分

5。

Let E be an embedding matrix, and let e_{1234} be a one-hot vector corresponding to word 1234. Then to get the embedding of word 1234, why don't we call $E*e_{1234}$ in Python?



It is computationally wasteful.

正确

\bigcirc	The correct formula is $E^Tst e_{1234}$.
	This doesn't handle unknown words (<unk>).</unk>
\bigcirc	None of the above: Calling the Python snippet as described above is fine.



6。

When learning word embeddings, we create an artificial task Natural Language aring of the solution of the solu

10/10 分 (100%)

测验, 10 个问题

on this artificial prediction task; the more important byproduct of this task is that we learn a useful set of word embeddings.

CITIDEA	um ₆ 5.
	True
正确	
	False
~	1/1分
the tar	word2vec algorithm, you estimate $P(t\mid c)$, where t is get word and c is a context word. How are t and c from the training set? Pick the best answer.
	\emph{c} and \emph{t} are chosen to be nearby words.
正确	
	\boldsymbol{c} is the sequence of all the words in the sentence before \boldsymbol{t} .
	c is the one word that comes immediately before t .

 \emph{c} is a sequence of several words immediately before

t .

Suppose you have a 10000 word vocabulary, and are learning Natural Language Processing Bed More he made defines uses 10/10 \$ (100%)

测验, 10 个问题

the following softmax function:

$$P(t \mid c) = rac{e_{ heta_t}^{ extit{Tec}}}{\sum_{t'=1}^{10000} e_{ heta_t'}^{ extit{Tec}}}$$

Which of these statements are correct? Check all that apply.

 $heta_t$ and e_c are both 500 dimensional vectors.

正确

 $heta_t$ and e_c are both 10000 dimensional vectors.

未选择的是正确的

 $heta_t$ and e_c are both trained with an optimization algorithm such as Adam or gradient descent.

正确

After training, we should expect $heta_t$ to be very close to e_c when t and c are the same word.

未选择的是正确的



1/1分

Suppose you have a 10000 word vocabulary, and are learning 500-dimensional word embeddings. The GloVe model minimizes this objective:

$$\min \sum_{i=1}^{10,000} \sum_{j=1}^{10,000} f(X_{ij}) (heta_i^T e_j + b_i + b_j' - log X_{ij})^2$$

Which of these statements are correct? Check all that apply.

 θ_i and e_j should be initialized to 0 at the beginning of training.

Natural Language Processing & Word Embeddings

10/10 分 (100%)

测验, 10 个问题

未选择的是正确的

 θ_i and e_j should be initialized randomly at the beginning of training.

正确

 X_{ij} is the number of times word i appears in the context of word j.

正确

igwedge The weighting function $f(.\,)$ must satisfy f(0)=0 .

正确

The weighting function helps prevent learning only from extremely common word pairs. It is not necessary that it satisfies this function.



1/1分

10_°

You have trained word embeddings using a text dataset of m_1 words. You are considering using these word embeddings for a language task, for which you have a separate labeled dataset of m_2 words. Keeping in mind that using word embeddings is a form of transfer learning, under which of these circumstance would you expect the word embeddings to be helpful?



 $m_1 >> m_2$

正确

 \bigcirc $m_1 \ll m_2$

Natural Language Processing & Word Embeddings

10/10 分 (100%)

测验, 10 个问题

