

<u>Unit 5 Reinforcement Learning (2</u>

Lecture 19: Applications: Natural

Course > weeks)

6. Word Embeddings

> <u>Language Processing</u>

### **Audit Access Expires May 11, 2020**

You lose all access to this course, including your progress, on May 11, 2020.

# 6. Word Embeddings Word Embeddings





Video

Download video file

**Transcripts** 

<u>Download SubRip (.srt) file</u>

Download Text (.txt) file

## **Understanding Word Embeddings**

1 point possible (graded)

Which of the following options is correct about word embeddings presented in the lecture.

The goal of word embeddings is to increase the sparsity of the encoded input word features

We would like similar words to have word embeddings that are far apart in order to minimize word sense disambiguation

One way to learn word embeddings is by maximizing cosine similarity between words with related meaning. ✓

To do a good job, word embeddings have to manually encoded by a natural language domain expert

#### **Solution:**

We would like to learn word embeddings that are much less sparse than one hot vector based encoding because reducing the sparsity of input features lowers the sample complexity (number of training examples required to do an accurate task) of the downstream text classification task.

In order to do the above, we should cluster the similar or related words together in the embedding dimension space. For instance, the words "dog" and "samoyed" must have similar embedding representations than "dog" and "lipstick".

Word embeddings are practically very useful because they can be learnt without any significant manual effort and they generalize well to completely new tasks.

Submit

You have used 0 of 2 attempts

**1** Answers are displayed within the problem

# Sentence Embeddings vs bag-of-words

1 point possible (graded)

Consider the following two sentences with very different meanings:

- (1) I ate pizza with my friend
- (2) I ate my friend with pizza

	Bag of words encoding approach wo	uld lead to identical	encodings for	both
t	hese sentences 🗸			

Bag of words encoding approach would be able to successfully differentiate
the above two sentences

Recurrent neural network (e.g., LSTMs) based approach would lead to
identical encodings for both these sentences

Recurrent neural network (e.g., LSTMs) based approach would be able to
successfully differentiate the above two sentences 🗸

#### **Solution:**

Bag of words approach sums up all the word embedddings in order to encode an input sentence. Hence, it cannot capture the ordering of these words within a sentence.

LSTM or Recurrent Neural network based approaches encode an input sentence into a context vector capturing more than just a summation of its constitutent parts together.

Submit

You have used 0 of 1 attempt

**1** Answers are displayed within the problem

## Discussion

**Hide Discussion** 

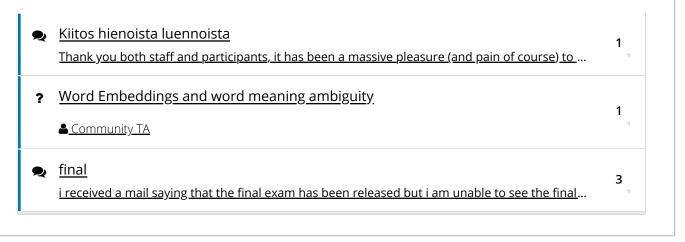
**Topic:** Unit 5 Reinforcement Learning (2 weeks) :Lecture 19: Applications: Natural Language Processing / 6. Word Embeddings

Add a Post

Show all posts by recent activity	
<ul> <li>Nice to see you all at the end of the lecture. Good job everyone!</li> <li>Thanks for you guys to go through the course with us in this difficult period. It is really a speci</li> <li>Staff</li> </ul>	28
thanks thanks professors barzilay, jakoola for your great lectures and wonderful team.enjoyed,thanks!	3
Thanks Thanks to professors, all of the staff, TAs!	2
Thank You!!  Thank you Profesors Barzilay, Jakoola and TA's for this amazing course.! This was tough at ti	2
Thanks [can't believe this is the last video in the course] Thanks a lot for our great lecturers and TAs I can't get enough of learning about these stuff	2
Thanks Thank you Professors Barzilay, Jakoola and TA's	29
? To which parts of the course she refer to ?  To which parts of the course she refer to when she speak of embedding words in a dense co  **Community TA*	1
? Final exam  Does the final exam include questions for this unit?	2

6. Word Embeddings | Lecture 19: Applications: ...

https://courses.edx.org/courses/course-v1:MITx+...



© All Rights Reserved

5 of 5 2020-05-09, 9:52 a.m.