Lecture3 Neural Networks

CS224N第三章前面主要介绍了神经网络方面的知识，这里就不做整理了。这章后面部分主要以**命名实体识别** Named Entity Recognition 来讲DNN在NLP领域内的应用。

Named Entity Recognition (NER)

命名实体识别是什么

学完这部分后回过来，我认为首先我们要理解实体是什么。

简单的理解，**实体可以被认为是某一个概念的实例。**例如“人名”是一个概念，或者说是实体类型，那么“吴天鹤”就是一种“人名”的实体。“时间”是一种实体类型，那么“国庆节”就是一种“时间”实体了。而**所谓实体识别，就是将你想要获取到的实体类型，从一句话里挑出来的过程。**

这里可以举个例子：
小明 在 北京大学 的 燕园 看了 中国男篮 的一场比赛
PER ORG LOC ORG

对于上面的例子，句子“小明在北京大学的燕园看了中国男篮 的一场比赛”，通过NER模型，将“小明 ”以PER，“北京大学”以ORG，“燕园”以LOC，“中国男篮”以ORG为类别分别挑了出来。

命名实体识别的数据标注方式

NER是一种序列标注问题，因此他们的数据标注方式也遵照序列标注问题的方式，主要是BIO和BIOES两种。这里直接介绍BIOES，明白了BIOES，BIO也就掌握了。

先列出来BIOES分别代表什么意思：

B，即Begin，表示开始

I，即Intermediate，表示中间

E，即End，表示结尾

S，即Single，表示单个字符

O，即Other，表示其他，用于标记无关字符

将“小明在北京大学的燕园看了中国男篮的一场比赛”这句话，进行标注，结果就是：

[B-PER，E-PER，O, B-ORG，I-ORG，I-ORG，E-ORG，O，B-LOC，E-LOC，O，O，B-ORG，I-ORG，I-ORG，E-ORG，O，O，O，O]

那么，换句话说，**NER的过程，就是根据输入的句子，预测出其标注序列的过程。**

命名实体识别的任务

- The task: **find** and **classify** names in text, for example:

  The European Commission [ORG] said on Thursday it disagreed with German [MISC] advice.

  Only France [LOC] and Britain [LOC] backed Fischler [PER] 's proposal .

  “What we have to be extremely careful of is how other countries are going to take Germany 's lead”, Welsh National Farmers ' Union [ORG] (NFU [ORG]) chairman John Lloyd Jones [PER] said on BBC [ORG] radio .

  其实对于NER问题的定义就是找到文本种的名字并对其进行分类。如上图中标黄色的字眼。

  找到地点类的名词France，Britain以及人名Fischler。

- **Possible purposes**
  - Tracking mentions of particular entities in documents
  - For question answering, answers are usually named entities
  - A lot of wanted information is really associations between named entities
  - The same techniques can be extended to other slot-filling classifications

Named Entity Recognition on word sequences

We predict entities by classifying words in context and then extracting entities as word subsequences.

Why might NER be hard

- **Hard to work out boundaries of entity**
  - 这里有一个简单的例子说明NER有时候很难区分Named Entity的边界

    Is the first entity "First National Bank" or "National Bank"

- **Hard to know if something is an entity**

  Is there a school called "Future School" or is it a future school?

- **Hard to know class of unknown/novel entity**

To find out more about Zig Ziglar and read features by other Creators Syndicate writers and

What class is "Zig Ziglar"? (A person.)

- **Entity class is ambiguous and depends on context**

where Larry Ellison and Charles Schwab can
live discreetly amongst wooded estates. And

"Charles Schwab" is PER not ORG here

Binary word window classification

鉴于同一个词在不同上下文可能是不同的Named Entity，一个思路是通过对该词在某一窗口内附近的词来对其进行分类（这里的类别是人名，地点，机构名等等）。

Window classification

Idea

Classify a word in its context window of neighboring words.

A simple way to classify a word in context might be to **average** the word vectors in a window and to classify the average vector.

- Problem: **that would loss position information.**

Named Entity Classification

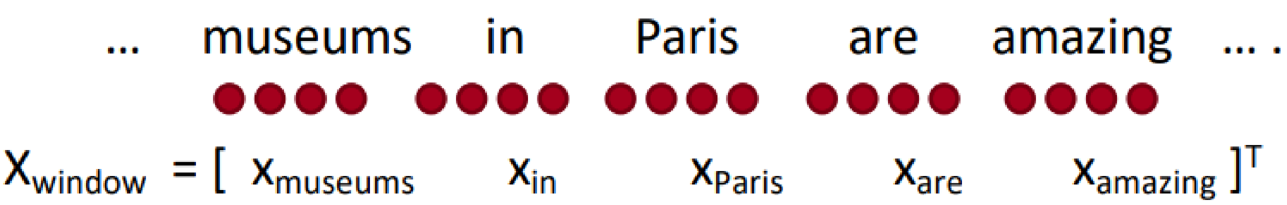
Person, Location, Organization, None

Window classification: Softmax

Train softmax classifier to classify a center word by taking **concatenation of word vectors surrounding it** in a window.

Example

Classify "Paris" in the context of this sentence with window length 2:



Resulting vector $x_{window} = x \in R^{5d}$, a column vectoe!

Simplest window classifier: Softmax

predicted model output probability

$$\boxed{\hat{y}_y} = p(y|x) = \frac{\exp(\boxed{W_y.x})}{\sum_{c=1}^{C} \exp(W_c.x)}$$

- With cross entropy error as before:

same

$$J(\theta) = \frac{1}{N} \sum_{i=1}^{N} -\log\left(\frac{e^{\boxed{f_{y_i}}}}{\sum_{c=1}^{C} e^{f_c}}\right)$$

Binary classification with unnormalized scores

- For our previous example:

  X_window = [x_museums x_in x_Paris x_are x_amazing]

- Assume we want to **classify whether the center word is a Location**
- Similar to word2vec, we will go over all positions in a corpus. But this time, it will be supervised and **only some positions should get a high score**.
- E.g., **the positions that have an actual NER Location in their center are "true" positions and get a high score**

Binary classification for NER Location

Example

Not all museums in Pairs are amazing.

这里只有一个真实的window就是Pairs在中心,而其余的windows我们称它们为"corrupt".

正确的window是: [museums in Pairs are amazing]

- "Corrupt" windows are easy to find and there are many: Any window whose center word isn't specifically labeled as NER location in our corpus

Neural Network Feed-forward Computation

Use neural activation a simply to give an unnormalized score

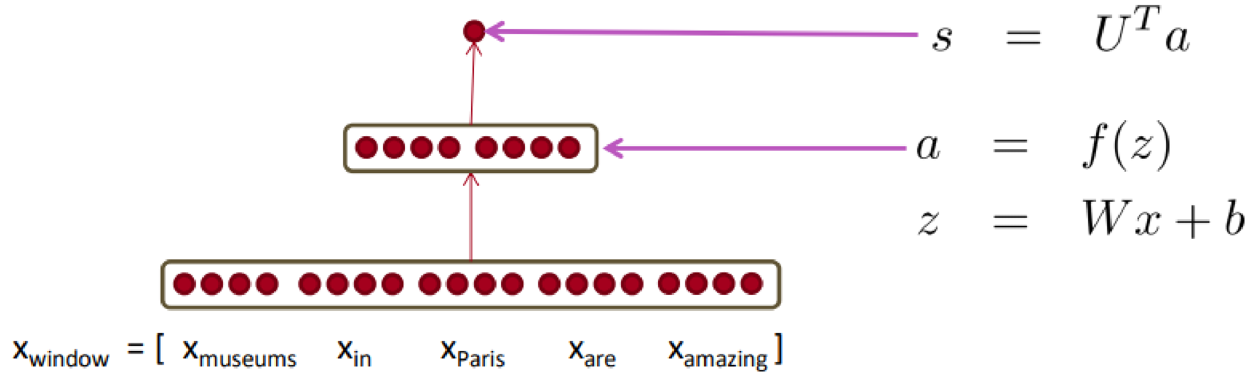
$score(x) = U^T a$

We compute a window's score with a **3-layer neural net**:

- s = *score*("museums in Paris are amazing”)

$$s = U^T f(Wx + b)$$

$$x \in \mathbb{R}^{20\times 1}, W \in \mathbb{R}^{8\times 20}, U \in \mathbb{R}^{8\times 1}$$



34

The middle layer learns **non-linear interactions** between the input word vectors.

The max-margin loss

Idea for training objective

Make true window's score larger and corrupt window's score lower (until they're good enough)

- $s = score(museums, in, Pairs, are, amazing)$
- $s_c = score(Not, all, museums, in, Pairs)$
- Minimize

  $J = max(0, 1 - s + s_c)$
- This is not differentiable but it is continuous: we can use SGD
- Each window with an NER location at its center should have a score +1 higher than any window without a location at its center

xxx |← 1 →| ooo