Bo Xia

Session 5B (room TU101): Semantics

2018年11月23日

Distributed Representation of Chinese Collocation

Let’s begin.

Hello, ladies and gentlemen. I’d like to start by introducing myself, my name is xiabo, student at Beijing Language and Culture University

I plan to speak about collocation representation, what I would like to do today is to illustrate an unsupervised algorithm for Collocation Representing in vector space using segmented corpora.

What I want my listeners to get out of my speech is deep neural networks or deep learning, and its corresponding field of natural language understanding. If there is one thing I’d like to get across to you today it is that sparse data is killing ML, which was the bottleneck in development of NLP application. For example, Sparse data can be low frequency of ngrams, or lack of sentence structure. To overcome words sparsity, formula of distributed representation is used to …, to overcome collocation sparsity, emmmm.

Why we have to represent collocation?

As an important role of language phenomenon, lexical collocations is still far from work in nature language processing such as word sense disambiguation, parsing and topic modeling

Due to rich information and features of linguistic collocation representation, distributed vector representations of collocations allow us to mine collocation features statistically and consider word combinations as a unit during parsing if necessary.

I have divided my presentation into Y parts

In the first part I give a few basic definitions.

1. Introduction 提出若干问题引起注意

WHY表示搭配

什么是搭配，搭配的来源

MWEs

如何表示关系，三元组

如何抽取关系，

WHY, WHAT, WHO, WHEN, WHERE, HOW

动机：抽取搭配关系，之后过滤出动宾关系

Is China Winning the *Trade War* ?

<https://www.newyorker.com/news/our-columnists/why-china-is-winning-the-trade-war>

1. 问题描述WHAT

By collocation, we mean a directed association of head word and dependency word that constructed by fully random combine in a sentence.

We provide computational linguistics an unsupervised algorithm for *Collocation Representing*

Our algorithm represents dependency words by dense vectors that are trained to predict contexts of the head word. We show that these vectors provide high performance for extracting collocation similarities in syntactic and semantic.

The quality of work results performs effective on our test set, which is measured in a verb-verb phrase collocation prediction task.

1. 主要思想：HOW
   1. 思想介绍

方法一，每个搭配训练一个向量，大小=搭配数，缺点

方法二，

“You shall know a word by the company it keeps” (Firth, 1957).

每个词训练一个包含搭配信息的向量，大小=词表大小 优点

* 1. 流程图：

1. 创新点
2. 实验
   1. 数据源
   2. 准确率
   3. DEMO展示
3. 反思和改进
4. 常见方法，本方法的具体实现

I plan to speak about TITLE AND SUBJECT

What I would like to do today is

to explain

to illustrate

to give my background information on

to outline

to have a look at

What I want my listeners to get out of my speech is

If there is one thing I’d like to get across to you today it is that

I have divided my presentation into Y parts

In the first part I give a few basic definitions.

In the next section I will explain

In part three, I am going to show

In the last part I would like to give a practical example..

There will be plenty of time at the end of my speech for a discussion

Now let us turn to point one

Let us now move on to the second part, which is, as I said earlier

There are three things we have to consider: one, two, three

Now let us look at the first aspect which is

First of all,

Content

Quantity

Sequencing your ideas

Keeping the audience’s attention

Signposting where you are

That’s all I would like to say about subject of part A, and now let us turn to

Now that we’ve seen … let us turn to …

Now let’s take an example

An example of this can be found

To illustrate this

Let’s see this through an example

For example,

For instance,

Let me rephrase that,

In other words

Another way of saying the same thing is

That is to say

What is very significant is

What is important to remember

I’d like to emphasize the fact that

I’d like to stress the importance of

What I tried to bring out

What we need to focus on

To summarize

To sum up,

Let me summarize by saying

So that concludes my overview

In conclusion

Briefly said

In short,

What I’ve tried to show in this part

To recap what we’ve seen so far

As I already said earlier

As we saw in part one

To repeat what I’ve said already

We will see this a little later on

This will be the subject of part 3

We will go into more detail on that later

I quote the word of

In the word of

According to

Here I’d like to quote

As Mr. X says in his book

There is a famous quotation that goes

As you all may well know

It is generally accepted that

As you are probably aware of

I’d like to summarize/sum up

At this stage I would like to run through/over the main points

So, as we have seen today

As I have tried to explain this morning BT finds itself in

In conclusion I would like to say that

My final comments concern

I would like to finish by reminding everyone that

I’d be happy to answer any questions

If there are any questions please feel free to ask

Thank you very much for your attention and if there are any suggestions or comments

Let’s look at current distribution of the market, as you can see

I’m going to show you now the most recent figures available

My next slide concerns the method by which