

# Assessing the Ability of LSTMs to Learn Syntax-Sensitive Dependencies

这里介绍一篇论文 : Assessing the Ability of LSTMs to Learn Syntax-Sensitive Dependencies

这个是论证 LSTMs 结构对于依存结构的层级性语法信息的抓取不可靠的论文.

## 1 实验设计

这里设计的实验是谓词单复数预测问题. 这个任务的难点就在于 (主语-动词) 关系的预测. 比如下文:

- 简单的例子 : The **key is** on the table (正确) | The **key are** on the table (错误)
- 复杂的例子 : Alluvial **soils** carried in the *floodwaters* **add** nutrients to the floodplains.

可以看到第二个句子中动词前面有两个名词 : soils 和 floodwaters. 一个是单数一个是复数, 这样的话就需要动词去判断谁是主语.

**重点 :** 这个里面, 真实主语和真实动词之间的夹杂的, 与**真实主语不同number**的名词 叫做 **agreement attractors (intervening nouns)**

## 2 模型设计

这里设计了两个模型.

正常模型 : 句子输入LSTM, 再经过一个逻辑回归器

Baseline : 只有名词组成的句子(即抽出其他词汇)输入LSTM, 再经过一个逻辑回归器(含有两个变种)

- common nouns only : 动词全为一般名词, 即 dogs, pipe 之类的
- all nouns : 加入了代词 (he) 和 专有名词 (France).

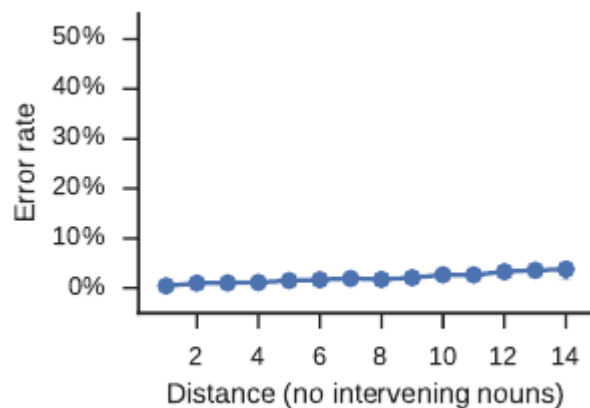
这里面还有两个很重要的调节参数, 就是:

- agreement attractors(intervening nouns)) 的数目
- 句子长度

## 3 实验结果

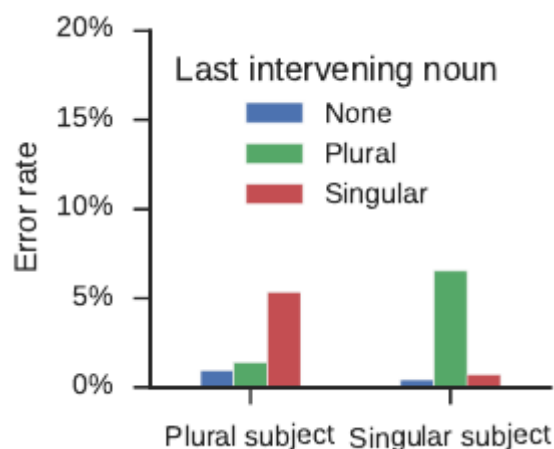
### 1) 随着距离的上升, 错误率提升

这个是针对正常模型.

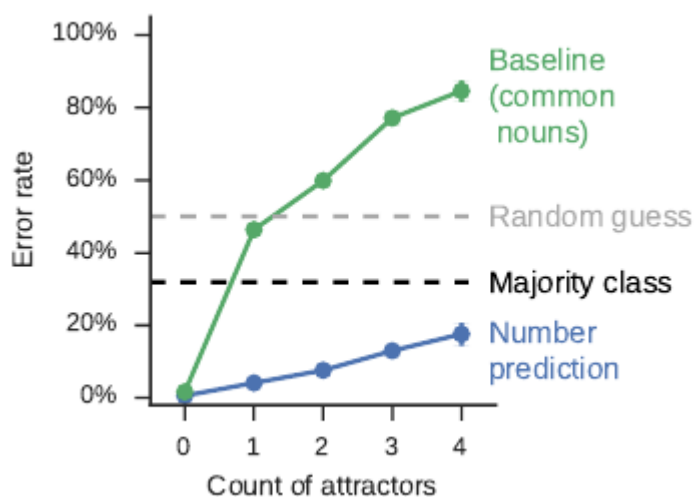


## 2) intervening nouns 对结果影响很大

这个也是针对正常模型



## 3) intervening nouns 数目的影响 与 hard cases 的影响



图形解释:

- 蓝色的线是 正常模型

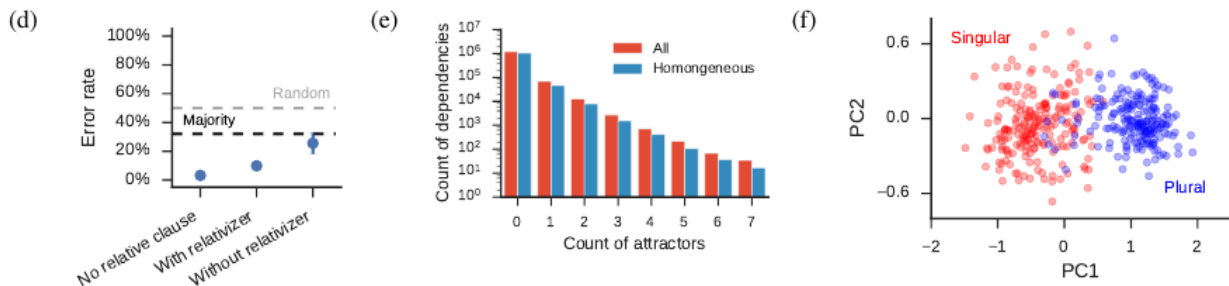
- 绿色的线是 baseline 模型

分析

- 可以看出, 随着intervening nouns数目的增长, performance是越来越差的, 但是对于蓝色的线来说还好
- 但是对于 baseline 的情况, 效果差完了
- 说明语法信息对于这个任务来说及其重要. 只看名词词汇是不行

#### 4) 其他的实验

有兴趣的自己看吧



#### 4 另外的实验

这里还对几个不同的任务进行了试验, 但是其最后任务还是求单复数:

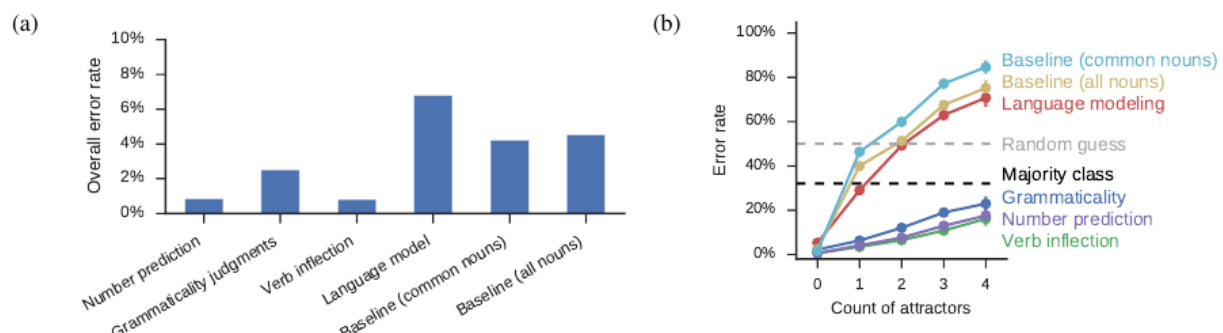
- Verb inflection : 给出动词的基本形式, 判断其单复数, 相比于之前的加上了其语义信息.
- Grammaticality judgments : 给出一个句子判断其语法正确与否
- Language modeling (LM) : 语言模型, 这里的求法见下面.

例子如下:

Training objective	Sample input	Training signal	Prediction task	Correct answer
Number prediction	<i>The keys to the cabinet</i>	PLURAL	SINGULAR/PLURAL?	PLURAL
Verb inflection	<i>The keys to the cabinet [is/are]</i>	PLURAL	SINGULAR/PLURAL?	PLURAL
Grammaticality	<i>The keys to the cabinet are here.</i>	GRAMMATICAL	GRAMMATICAL/UNGRAMMATICAL?	GRAMMATICAL
Language model	<i>The keys to the cabinet</i>	are	$P(are) > P(is)?$	True

Table 1: Examples of the four training objectives and corresponding prediction tasks.

实验结果:



## 5 结论

- LSTMs were able to learn to perform the verb number agreement task in most cases
- 但是对于长的句子和难的句子会有一些错误率的提升(这个结论只看a图就好了)
- 对于一个 structure-sensitive dependencies 重要的任务, 单单靠语言模型是不够的, a joint training objective是非常重要的. 这里就是说, 也就是说, 我们可以用多个任务来学习同一份词向量. 这样才可以在语言模型中加入语法信息.(个人感觉这里应该说的是 multi-task)