AI Vision Lab
Proposal

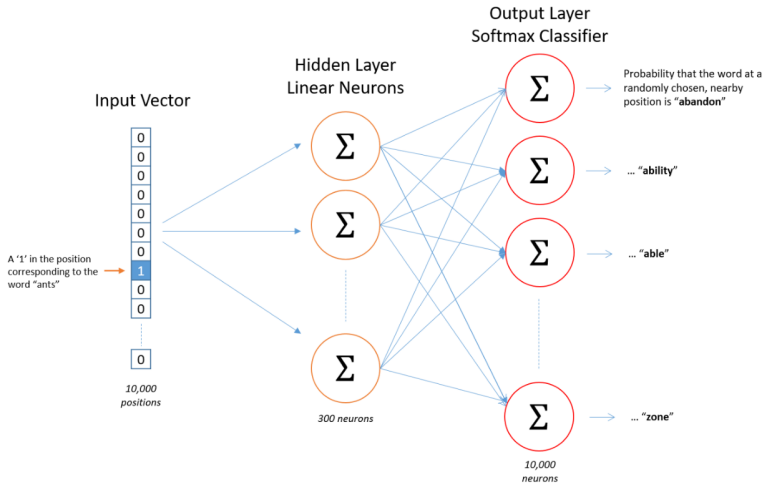
Sarah Khan

June 7, 2020

1 Objective

When answering the questions where we have to provide spatial information we tend to reply with reference to stabler objects. For example, We are in Munich rather than We are in car. Another example, The books are on the shelf rather than Shelf is under the book. Hence to be able to correctly express we need to posses Common Sense Knowledge of Stability of objects. This is exactly our objective for this lab:From a given text corpus learn Common Sense Knowledge of Stability.

Figure 1: SKIPGRAM model)



2 Proposed Solution

In SKIPGRAM Model we have architecture as shown in Fig 1 [1], where the training objective is to learn vector representation or word embedding of a word that is

good at predicting similar words to the given word. The proposal states that an extra dimension should be added to the vector representation which could be visualised as radius in space. This radius could represent the stability of the object and the larger the magnitude of vector the stabler the object. For obtaining the aforementioned change, we could add an extra hidden unit and an extra output unit which would then only be learned whenever we want to learn stability. The added output unit will be performing an exponential of the value rather than applying softmax,

$$e^{\beta x} \tag{1}$$

so that the radius is never negative. Also, for controlling the magnitude of this radius we can have a hyperparameter beta multiplied with the value. For leveraging the stability knowledge from text corpus we will use dependency graph as proposed by Levy, Omer Goldberg, Yoav. (2014) but in addition we will be using the preposition dependency to learn stability. So, as proposed by Levy, Omer Goldberg, Yoav. (2014) in case of preposition they collapse preposition and object into one and then combine the respective word and object as word,context pair. For example, a sentence is Australian scientist discovers star with telescope. Here we have prep(discovers,with) and pobj(with,telescope), both stated pairs are combined to give pair prep_with(discovers,telescope) along with other word-contexts pair obtained from dependency graph. This method would not work for our case because we do not need relationship between verb and noun rather we want relationship with a noun and a noun. So, apart from keeping these pairs, we need to form our own pairs to get the stability relationship. One such example is by combining nsubj,prep and pobj. Take the example sentence, The cup is on the table; with the original proposed method we would have the following pairs:
det(cup,the)
nsubj(is,cup)
prep(on,is)
det(table,the)
pobj(table,on)
prep_on(table,is)
but after combing nsubj,prep and pobj(only when this is noun) , we should have a new pair (table,cup).

Another example which also illustrates why pobj should be noun only. A man in a red shirt , here both red and shirt are pobj of preposition in and we only need to consider shirt for our case.

The objective function used in SKIPGRAM model was

$$argmax_{v_w,v_c}(\Sigma_{x,c\in D}\ log\ \sigma(v_c * v_w) + \Sigma_{x,c\in D'}\ log\ \sigma(-v_c * v_w)) \tag{2}$$

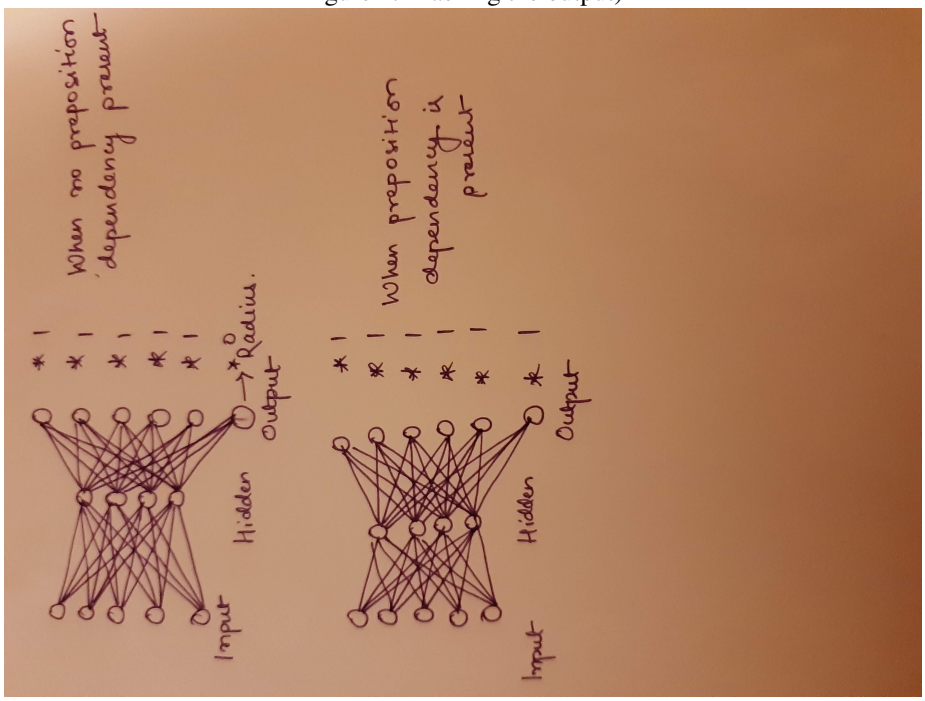
where each word $w \in W$ is associated with a vector $v_w \in R^d$ and similarly each context $c \in C$ is represented as a vector $v_c \in R^d$, where W is the words vocabulary,C is the contexts vocabulary, and d is the embedding dimensionality. D is positive dataset i.e. dataset consisting of actual word context pairs while D' is dataset for negative sampling which was constructed as proposed by Mikolov et al. To this existing objective function we can add an extra term $(r_{v_w} - r_{v_c})^2$ where r is the radius which has been obtained after exponential as stated beforehand and hence, the objective function then

becomes

$$argmax_{v_w,v_c}(\Sigma_{x,c\in D}\ log\ \sigma(v_c * v_w) + \Sigma_{x,c\in D'}\ log\ \sigma(-v_c * v_w)) + (r_{v_w} - r_{v_c})^2 \quad (3)$$

The radius is always positive as obtained from eq.(1) and for backpropagation we will have $2 * (r_{v_w} - r_{v_c})$. This tries to maximise the radius differences and at the same time maintain that radius of word is greater than that of its context. Backpropagation for learning the radius will only be performed when there is a prepositional dependency as stated above encountered. To endure that we could mask the output as explained in image below.

Figure 2: Masking the output)



In Mikolov etal. negative sampling is for each (w,c) $\in$ D id by constructing n samples (w,c1),...,(w,cn), where n is a hyperparameter and each cj is drawn according to its unigram distribution raised to the 3/4 power. Since stability is only between nouns we could have few part of negative sampling contruction by including cj as only nouns.

3 Implementation

Preprocessing the data: Get the word - context tuples including the dependencies. For the combination of nsubj,prep and pobj to obtain stability tuple we can introduce a dependency tag names stability. So, we would have a tuple like (is,cup , nsubj), (table,is,prep_on), (table,cup, stability) etc.

References

[1] Laura Lorenz. Forward propagation: Building a skip-gram net from the ground up.

[1]