April 24, 2020

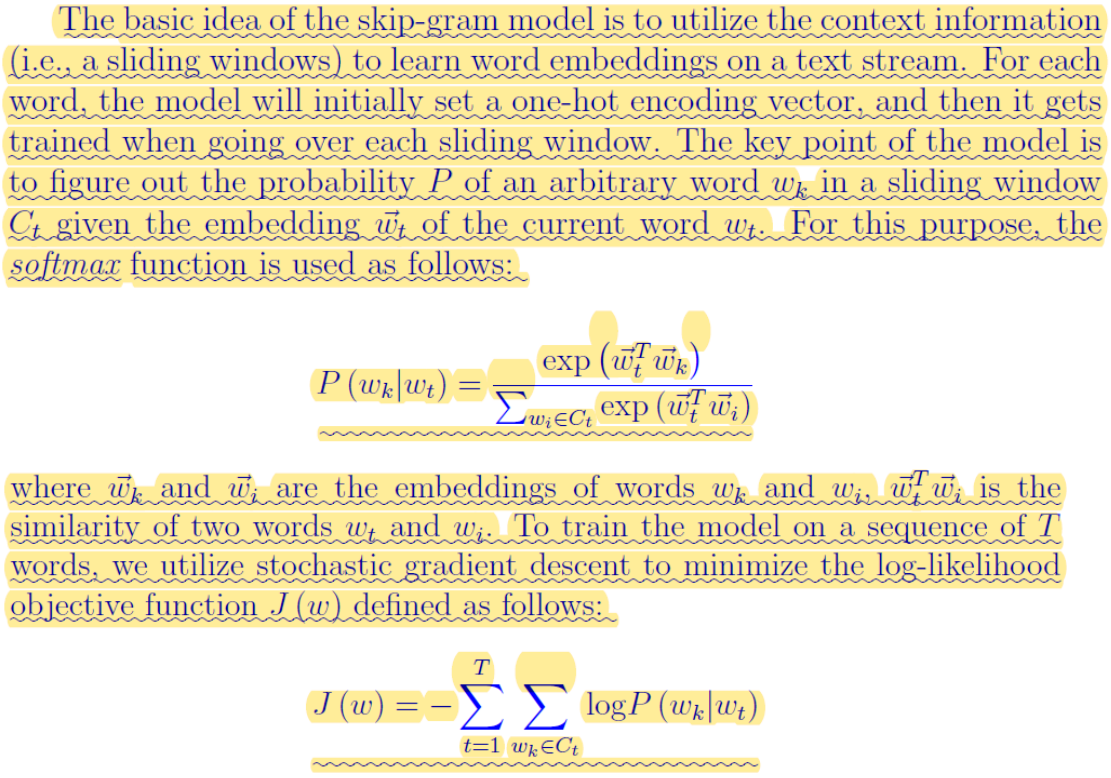
Review of the revised version of paper
Title: A Deep Learning Approach to Binary Code Similarity Detection
By Donghai Tian, Xiaoqui Jia, Rui Ma, Shuke Liu, Wenjing Liu, Changzhen Hu

Manuscript Number: **ESWA-D-05731R1**

I have just checked the way the Authors addressed my concerns/comments of February 8.

A) My Comment 1. of February 8 was: There is a problem in understanding of the method proposed in details, since it is presented rather in a descriptive way. More detailed mathematical description/formulation should make the paper better readable/clear.

a) My Comment 1 (of April 24) to the Authors response.
Indeed, in Section 2.3 the Authors have added a paragraph:



The basic idea of the skip-gram model is to utilize the context information (i.e., a sliding windows) to learn word embeddings on a text stream. For each word, the model will initially set a one-hot encoding vector, and then it gets trained when going over each sliding window. The key point of the model is to figure out the probability P of an arbitrary word w_k in a sliding window C_t given the embedding $\vec{w}_t$ of the current word w_t. For this purpose, the *softmax* function is used as follows:

$$P\left(w_k|w_t\right) = \frac{\exp\left(\vec{w}_t^T\vec{w}_k\right)}{\sum_{w_i\in C_t}\exp\left(\vec{w}_t^T\vec{w}_i\right)}$$

where $\vec{w}_k$ and $\vec{w}_i$ are the embeddings of words w_k and w_i. $\vec{w}_t^T\vec{w}_i$ is the similarity of two words w_t and w_i. To train the model on a sequence of T words, we utilize stochastic gradient descent to minimize the log-likelihood objective function $J\left(w\right)$ defined as follows:

$$J\left(w\right) = -\sum_{t=1}^{T}\sum_{w_k\in C_t}\log P\left(w_k|w_t\right)$$

But, to be more precise, the Authors addressed my concern not precisely/adequately. There should be still included element of which space/set is arbitrary word w_k ? Is this a sequence of bits, or of other symbols ? I pointed this in my Review of February 8.

What mathematical objects are the embedding $\vec{w}_t$? The same question is/arises to sliding window C_t ?

<u>Typically, one can use symbol “element of”, for example: $w \in Z^L$, Z={0,1}, L=5.</u>

b) Similar situation is in Section 4. Again, there is still no information about $(\vec{i}_1, \ldots, \vec{i}_n)$, $(\vec{\iota}_1, \ldots, \vec{\iota}_m)$, c_t, i_t, f_t, o_t, and ***b***. Elements of which spaces/sets these objects are ?

c) In Section 5 there is the same situation.
The objects/symbols used in this formula

$$A = \sigma\left(\mathrm{Conv}\left(\vec{X}\right)_{W_1, b_1}\right)$$

should be associated to some mathematically clear sets/spaces.

The same comments are to formulas:

$$O = \sigma\left(W_2 \cdot \widetilde{A} + b_2\right)$$

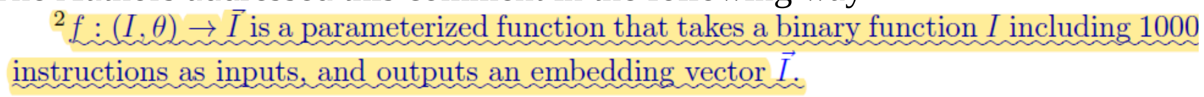
$$H = LSTM\left(O\right)_{W_3, b_3}$$

When one uses some symbols, one expects that these symbols will be associated to some mathematical well defined objects (real numbers, Cartesian product of real numbers, sequences of symbols from well-defined alphabet etc.).
Otherwise, the formulas look like a heuristic one.

B) The above my Comments address also the response of the Authors to my Major Comment2 (of February 8).

C) Concerning my Comment3 of February 8.

The Authors addressed this comment in the following way



[2] $f : (I, \theta) \rightarrow \vec{I}$ is a parameterized function that takes a binary function I including 1000 instructions as inputs, and outputs an embedding vector $\vec{I}$.

Again, I would replace the question: elements of which sets are I, θ, and $\vec{I}$??

D) Concerning my Comment4 of February 8. "Figure Captions should be more informative/detailed."

The Authors changed/completed the Caption in the following way:

NEW

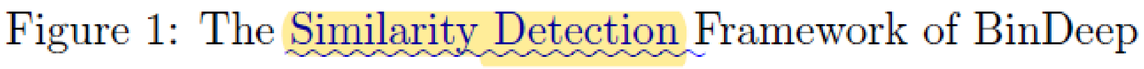


Figure 1: The Similarity Detection Framework of BinDeep

OLD

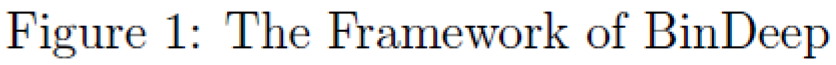


Figure 1: The Framework of BinDeep

===

NEW

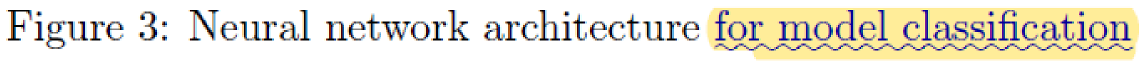


Figure 3: Neural network architecture for model classification

OLD

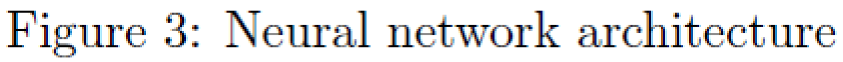


Figure 3: Neural network architecture

===

NEW

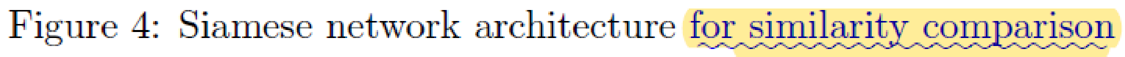


Figure 4: Siamese network architecture for similarity comparison

OLD

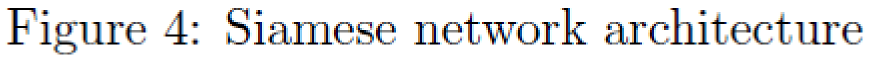


Figure 4: Siamese network architecture

===

When I was writing this comment I was thinking about including more detailed information. For example:
The input to the network are …
The output is …

etc.

At this moment the Authors have been added no additional more detailed information to the Figures Captions. They added a general information which in fact is obvious. The readers will not know no any additional information from the new Figure Captions.

FINAL COMMENTS

In my opinion, the Authors revised the paper not adequately/satisfactory to my comments. I do not recommend the new version of paper for publication. **Major revision in my opinion is still necessary.**