Deep Learning Theoretical Tasks for Lab 3

Authors: Axel Palmqvist, Wilmer Wohlén, Oscar Fahlén, Elias Falk

Task 3.1.1 - Explain the pros and cons of utilising Concatenation for combining embeddings

Concatenation is very intuitive and easy to understand. However, implementing it can be difficult to achieve due to it being a more complex operation when compared to simpler operations (taking the sum, averaging, taking the median, etc.).
Concatenation does not lose any information but does however create a high dimensionality. One can imagine a resulting vector after a concatenation being very large. Of course all information has been preserved but now there is instead a very large vector to handle instead.

Task 3.1.2 - Explain the pros and cons of utilising Addition for combining embeddings

Similar to concatenation, addition is very intuitive and easy to understand. Even more so than concatenation. Implementation is also easier to implement than concatenation but it runs the risk of losing information. The original information pertained to the added terms can be lost in the operation.
When it comes to vectors, addition has a requirement that both vectors must be of the same size. That is a limitation.

Task 3.1.3 - Explain the pros and cons of utilising Multiplication for combining embeddings

Similar to concatenation and addition, multiplication is intuitive and relatively easy to understand. Personally, we find multiplication the hardest to intuitively understand among the three operations mentioned so far. This operation also runs the risk of losing information. The original information pertained to the multiplied factors can be lost in the operation. Similar to addition, both embeddings need to be of the same size.

Task 3.1.4 - Explain the pros and cons of utilising Attention for combining embeddings

This operation can be challenging to understand but there are methods of making it more intuitive. As the name suggests, models that utilize attention can focus on parts of data that in certain contexts need to have greater impact. The operation does this by computing attention weights.
It is a very computationally expensive operation especially when compared to the simple operations mentioned above. It is also difficult to implement.
So all in all, this operation is expensive and difficult to implement but can really capture complex patterns and works much better with more complex data than the previously mentioned operations.

Task 3.1.5 - Explain the pros and cons of utilising Difference for combining embeddings

Similar to concatenation, addition and multiplication, difference, or subtraction as it is sometimes called, is intuitive and easy to understand.
It is very similar to addition in many ways. For instance, implementation is easy to implement but it runs the risk of losing information in a similar fashion. It also requires elements of terms of equal size.
It differs from addition and multiplication in the sense that this operation is not commutative, meaning the order of the elements in the operation matters. Another interesting thing about the difference operation is that the resulting element can be negative even if all terms are positive. This is not the case with the other addition, concatenation and multiplication operations. This can be an unwanted consequence in some contexts.