Weekly Report — (15 Mar, 2019)

Jiaru Zhang
AISIG
Shanghai Jiaotong University
800 Dongchuan Road, Shanghai

Abstract

1. I propose a new tensor decomposition method called Cylinder Decomposition. It is a generalization of Ring decomposition and CP decomposition. It is named as "cylinder" because it is like a stack of rings.
2. The performance of normal MBFN on Reuters dataset reaches 35 acc. as a multi-classification problem over 46 classes rather than about 10%. However, there still remains serious problems in the model.
3. I am despairing to find an error in the paper submitted to IJCAI. The theory descriped in the paper is a little different from what I used in experiments. More seriously, Sec. 3.2 and Sec. 3.4 of the paper is self-contradictory. I also found some expressions are not suitable.
4. To explain the tensor-based feature fusion methods theoretically, I read some papers. But they have nothing to do with the input features.
5. I show some of my thoughts and future plan.

1 Cylinder Decomposition

The CP decomposition is to factorize a tensor $\mathcal{W} \in \mathbb{R}^{A_1 \times A_2 \times \cdots \times A_n}$ into a sum of component rank-one tensors $\mathbf{U}_i^r \in \mathbb{R}^{1 \times A_i}$:

$$\mathcal{W} = \sum_{r=1}^{R} \left(\bigotimes_{i=1}^{n} \mathbf{U}_i^r \right). \tag{1}$$

The Ring decomposition is to express a tensor $\mathcal{W} \in \mathbb{R}^{A_1 \times A_2 \times \cdots \times A_n}$ into the product of a seires of three-order tensors $\mathcal{U}_i \in \mathbb{R}^{R_{i-1} \times A_i \times R_i}$:

$$\mathcal{W} = \prod_{i=1}^{n} \mathcal{U}_i. \tag{2}$$

If we choose all R (rank) in CP decomposition and Ring decomposition to 1, they will be equivalent. Motivated by this, I propose Cylinder decomposition, which expresses the tensor as a sum of Ring tensors:

$$\mathcal{W} = \sum_{r=1}^{R} \prod_{i=1}^{n} \mathcal{U}_i^r. \tag{3}$$

It is a stack of ring tensors so it is named as "cylinder". It can be also used as a feature fusion method just like Block decomposition.

2 MBFN on Reuters

I updated the codes of MBFN on Reuters and adjusted many hyper-parameters and settings. Now the acc reaches over 35% rather than original about 10%. However, the rank-one version can reach over 80%. So there are still many problems remained.

3 About the IJCAI paper

When I updated the MBFN on Reuters dataset, I found some discrepancies in the submitted paper. The parameters of $\mathbf{U}_i$ and $\mathcal{W}$ both have a dimension C in the experiments. However, in Sec. 3.2, only $\mathcal{S}$ has dimension C and in Sec. 3.4, only $\mathbf{U}_i$ has dimension C. It is indeed a big discrepency even though every independent section is OK.

Moreover, I found some expressions are not suitable because it is difficult to understand by others. For example, to show the product of matrix $A \in \mathbb{R}^{R_1 \times R_2}$ and matrix $B \in \mathbb{R}^{R_1 \times R_3}$ can be another matrix $C \in \mathbb{R}^{R_1 \times R_2 R_3}$, I should use an element-wise expression:

$$C(i, j_1 j_2) = A(i, j_1) B(i, j_2). \tag{4}$$

4 Paper Reading

To explain the performance of tensor-based feature fusion methods, I read some theoretical papers [1,2,3] explaining the expressive power of neural networks. They are useful and give me a lot of inspiration. However, all of them paid attention on the neural network structure itself rather than the effect of the input features.

5 Thoughts and Future Plan

Tensor decomposition is to use some simple tensors to express a complex tensor. It is relatively easy to propose novel tensor decomposition methods and use them in feature fusion. However, the theory to explain the performance difference of methods is poor. Moreover, the experimental improvement is indeed subtle.

Here is the future plan:

- Continue to focus on the MBFN on Reuters dataset.
- Read more relevant papers about the explanation neural networks.

References

[1] Raghu, Maithra . On the Expressive Power of Deep Neural Networks. In ICML, 2016.

[2] Guido Montu far etal. On the Number of Linear Regions of Deep Neural Networks. In NeurIPS, 2014.

[3] Nadav Cohen etal. On the Expressive Power of Deep Learning: A Tensor Analysis. JMLR: Workshop and Conference Proceedings vol 49:131, 2016.