**Paper: Layer-wise Relevance Propagation for Deep Neural Network Architectures**

**Alexander Binder, Sebastian Bach, Gregoire Montavon, Klaus-Robert Muller, and Wojciech Samek**

**0. Layer-wise Relevance Propagation**

|  |
| --- |
| Deep Neural Network은 다음과 같은 feed-forward graph이며, 각 neuron은 다음과 같은 함수를 이용한다.   * Deep Neural Network는 **많은 뉴런의 연결과 알고리즘(error backpropagation)**으로부터 complexity를 유도한다. * Neural Network의 **출력은 이 뉴런들을 feed-forward pass에서 계산하여 도출**된다.   또한, 이와 같은 그래프 구조는 **network의 output에서 relevance f(x)를 pixel-wise relevance score 로 재분배**하는 데 사용할 수 있다. 이때 사용되는 규칙은 다음과 같다.  B**ackward pass에서 이 규칙을 이용하면 relevance map (heatmap)**이 생성되는데, 이것은 다음과 같은 conservation property를 만족시킨다. -  **이 decomposition 알고리즘은 Layer-wise Relevance Propagation (LRP)** 이라고 한다. |

**1. and**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (2)에서의 Naïve propagation rule에서 다음과 같은 2개의 LRP 알고리즘이 있다.  **<>**   * 여기서 **이면 conservation idea는 더 좋은 numerical property를 얻기 위해 relaxate**된다.   **<>**   * **Relevance propagation 방정식이 conservative layer-wise**하게 하기 위하여 이 되게 한다.   지금까지 소개한 3가지 decomposition rule의 성질은 다음과 같다.   |  |  |  |  | | --- | --- | --- | --- | |  | naive () | -variant | -variant | | Numerically stable | NO | YES | YES | | Linear mapping과 일치 | YES | YES | NO | | Relevance 보존 | YES | NO | YES | |

**Source:** <https://s3.amazonaws.com/academia.edu.documents/52509826/BinICISA16.pdf>