Transformer(구글 attention is all you need)

Seq to Seq(인코더 디코더) attention 구조만을 사용, 자연어 처리

자연어처리 뿐 아니라 컴퓨터 비전, 음성처리

PTM으로 사용되었을 때 Downstream Task에서 SOTA를 보이고 있음(최종적인 task/state of the art)

최근, PTM모델을 fine tuning하여 높은 성능

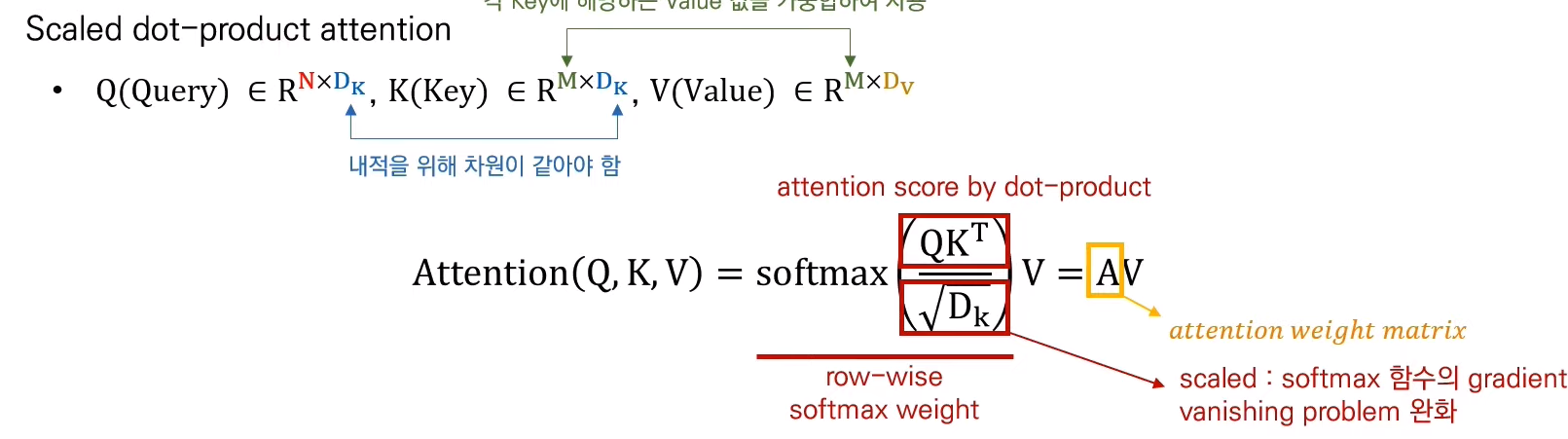
1. Model efficiency : self attention 모듈 비효율성 개선
2. Model Generaliation : 유연한 구조 -> input data의 structural bias에 대해서 최소한의 가정만을 하고 있어 적은 데이터 셋으로 강건한 모델 학습에 어려움이 있음. 개선….
3. Specific한 down stream

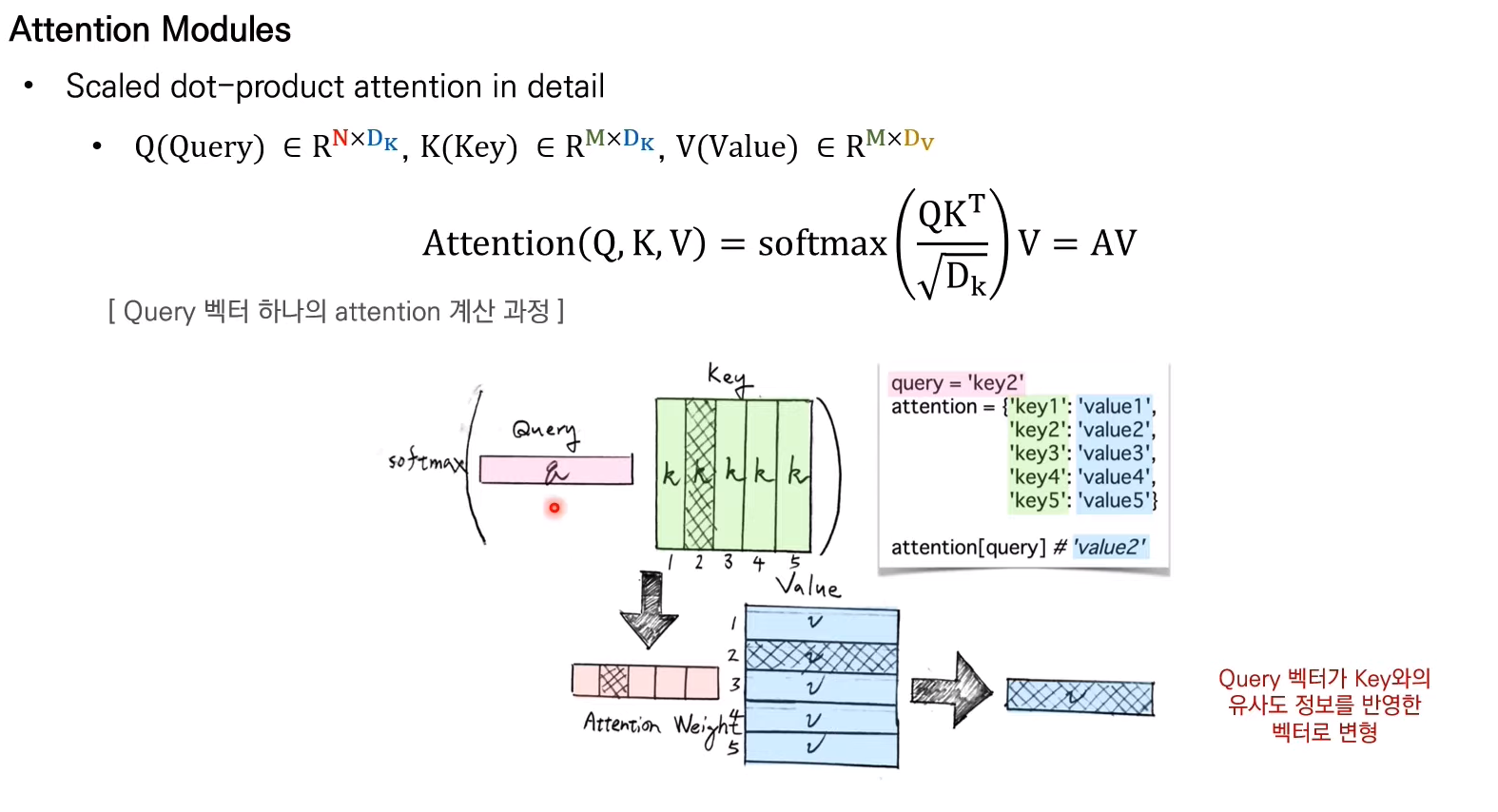
Vanilla Transformer 기본…

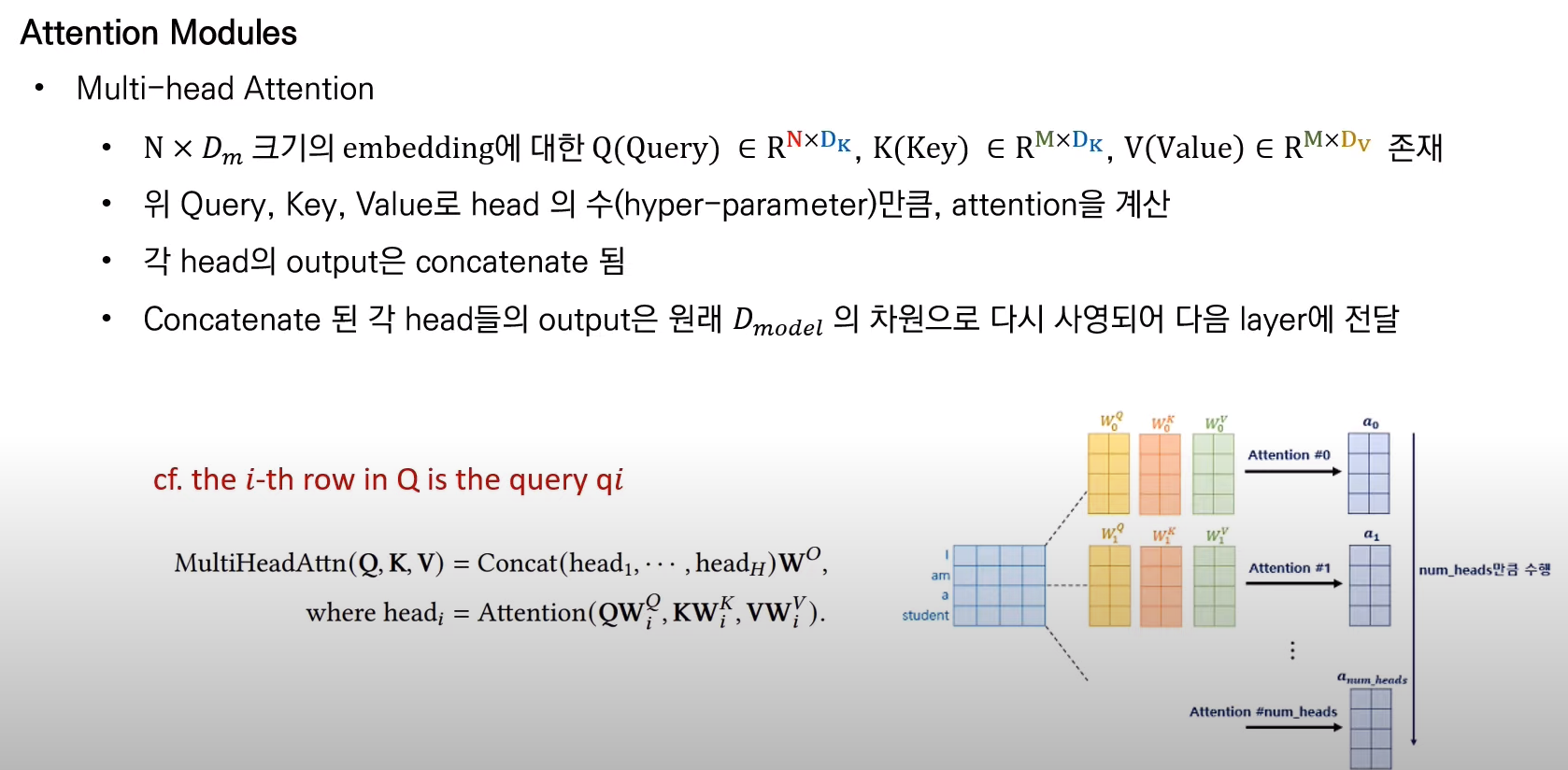
1. Seq2Seq
2. 각 encoder와 decoder는 L개의 동일한 Block이 Stack된 형태
3. Multi head self sttention, position wise feed forward net, residual connection, layer normalizqtion
4. 기본적인 encoder black, cross attention 모듈을 multi와 FFN 사이에 추가, 마스크사용

Attention Module

Query : 현재의 hidden state값, Key : hidden state과 영향을 주고 받는 값, value : query와 key의 가중합하여 query에 반영하게 되는데 가중합의 대상으로 사용할 값







다양한 특징 학습