

Source:

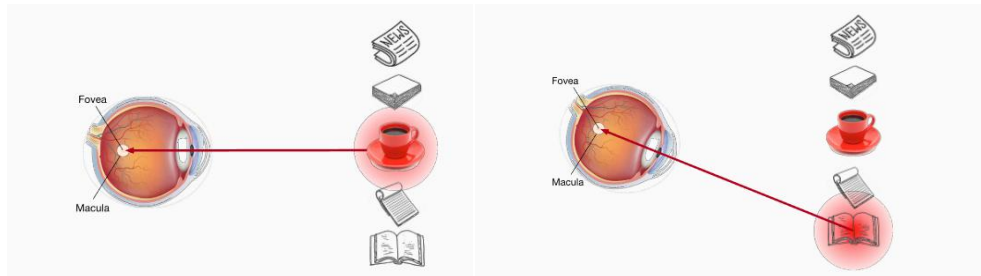
https://classic.d2l.ai/chapter_attention-mechanisms/attention-cues.html

https://en.wikipedia.org/wiki/Kernel_regression

Attention is all you need

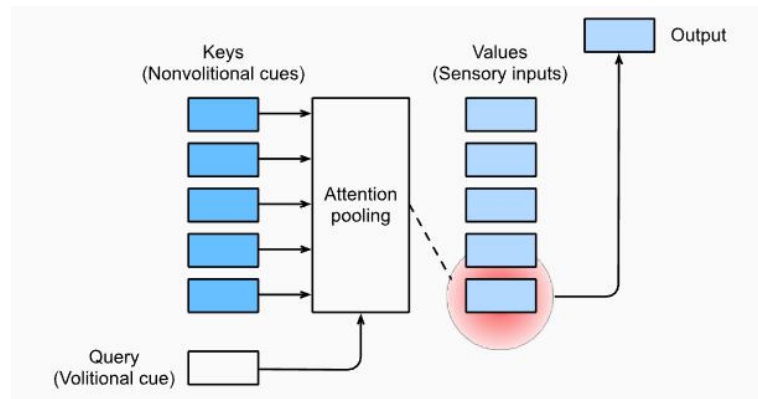
- attention mechanism 直观理解
 - 心理学 { 随意线索 - volitional clues \rightarrow 选注意点
不..... - non-volitional clues
 - 显式考虑 随意线索 \rightarrow query
 - 环境输入 key-value
 - 历史: ① 非参注意力池化层, Nadaraya-Watson 核回归
② 参量化注意力机制 (权重设计)
- 注意力函数, query 和 key 相似度
$$f(x) = \sum_i \alpha(x, x_i) y_i = \sum_{i=1}^n \text{softmax}(-\frac{1}{2}(x-x_i)^2) y_i$$
 - 拓展到高维
 - α 的设计?
 - additive attention $\alpha(k, q) = v^T \tanh(w_k k + w_q q)$
 - scaled dot-product attention $\alpha(q, k_i) = \langle q, k_i \rangle / \sqrt{d}$
- 使用 attention mechanism 的 seq2seq
 - 动机
 - 加入 attention mechanism
- 自注意力 (可略?)
 - 和 CNN, RNN 对比.
 - 位置编码.
- Transformer - only based on attention mechanism.
 - 多头注意力
 - masked 多头注意力

- Attention mechanism 的直观理解
- 心理学：分析两种线索选注意点
 - ◆ 随意线索 volitional cues
 - ◆ 不随意线索 non-volitional cues



https://classic.d2l.ai/chapter_attention-mechanisms/attention-cues.html

- 显式考虑 volitional clues
 - ◆ 随意线索认为是 Query
 - ◆ key, value 认为是环境输入
 - ◆ Key 和 query 做 attention pooling（类似卷积，但有不同的规则）



https://classic.d2l.ai/chapter_attention-mechanisms/attention-cues.html

- 历史：
 - ◆ 非参注意力池化层 Nadaraya-Waston 核回归
https://en.wikipedia.org/wiki/Kernel_regression
 - ◆ 参数化注意力机制（what about 权重设计？）