

Language models are unsupervised multitask learners

OpenCV 2019

提出 Model: GPT-2 (GPT model 的延伸 \Rightarrow params $\times 10$
dataset size $\times 10$)

Dataset: 自行构建，目的: as large and diverse as possible

- 亮点:
- ① Model 足够大 1.5B 参数量 \rightarrow performance - #params
 - ② 计算资源足够多 \Rightarrow 毕竟是 OpenCV
 - ③ Scalability 足够强 \Rightarrow 面对一系列 zero-shot task

\downarrow
四个任务 performance 圈

e.g. Machine Translation

Summarization

Reading Comprehension

(Q&A)

表现均优于 base line, 达到 state-of-art

WebText

- ④ Dataset 的获取:
(i) 利用 web scrape 从 Reddit 平台上
的 out bound links 获取质量较可靠的 web pages \Rightarrow
用户 share 的这些 links 多是其觉得有质量的, 因此相当于
借力进行了 human filtering, 得到优质的数据集.

(ii) 使用 Dragnet + Github上 - content extractor

提取 HTML response 中的 text 信息 \Rightarrow

extractor 架构图

(iii) Remove all Wiki documents \Rightarrow common data
source, 会使训练数据重叠, 不利于此后对 test
evaluation tasks 的分析.

⑤全文阐述最多的改变即解除此前 NLP 领域对 model, dataset 等的一系列限制, 给了 GPT-2 model 很大的自由度 \Rightarrow 避免出现 sub-optimal allocation 和压缩 model capacity.

\Rightarrow 也为 GPT-2 scalability 如此强大的原因。

- without preprocessing \Rightarrow BPE 合并相近的 word slots
- without task-specific restriction on test evaluation tasks
- without restriction on dataset construction
- without restriction on target application tasks

Survey on Explainability Recommendation

Task: 1. Survey chapter 5 ✓ ⇒ Explainable Recommendation 领域
全局性已解

2. Explainability model 有无什么现成的 dataset 可以做的 ✓ ⇒
具体的代表性 papers 的情况

Task 1:

(i)

Explainable Recommendation field 有几类应用领域？

(ii) Explainable Recommendation 对每个领域的 作用是什么？

(iii) 每个应用领域 如何选择 代表性的 Paper? ━━ # references → more
dataset → different

(iv) 每个应用领域的 代表性 Paper 用什么方法做 recommendation?

━ 用什么方法做 explanation?

Task 2:

(i) Paper 使用了具体什么 Explainable Recommendation model?

(ii) Paper 使用的数据集 来源? 形式是什么样的?

(iii) Paper 的 model 是如何利用数据集的?

它为何 不用其它数据集? 这个数据集能给此

paper 带来什么好处?

→ 如何获取内容?
是否经过预处理?

Explainable Recommendation

Categories:

Application

I. E-commerce recommendation

II. Place Of Interest recommendation

III. Social recommendation

IV. Multi-media recommendation

V. New fields

Science paper recommendation

Medical medicine recommendation

Educational recommendation

I. ① 提高建议说服力 ② 有助于用户做出有效知情的决定

II. ① 帮助用户节约时间、犯错成本 ② 更好规划旅行路线
③ 提高建议说服力

Effects on
different fields

III. ① 提高用户对推荐结果的置信度 (trustworthiness)

② 帮助用户快速识别有用信息，节约时间

③ 帮助用户在社会环境中识别虚假信息

IV

① 节约用户时间 → 不用看不相关的内容，快速找到自己想要的内容

② 避免不必要的多媒体类数据传输