

BAI

智源学者成果展示——智源青年科学家

作者 兰艳艳、郭嘉丰（中科院计算所）

2020年6月

项目负责人：兰艳艳、郭嘉丰

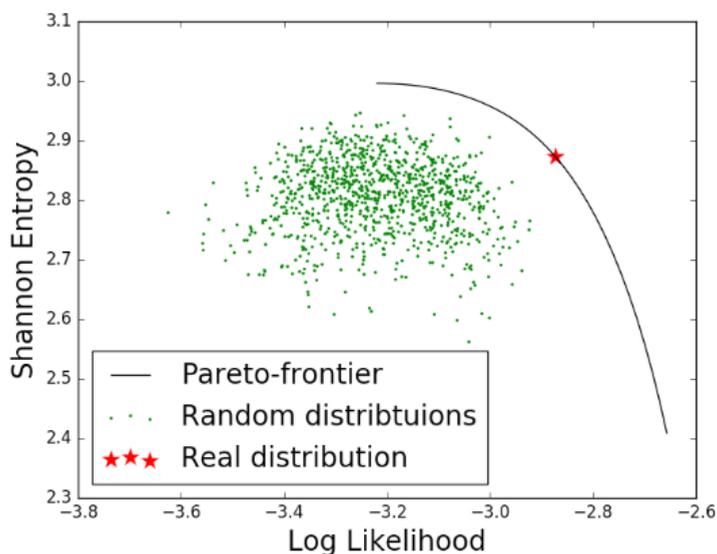
成果名称：

中文：论文本生成中质量 / 多样性评价与分布拟合目标之间的关系

英文：On the Relation between Quality–Diversity Evaluation and Distribution–Fitting Goal in Text Generation

成果简介：

文本生成模型的目标是拟合文本数据的概率分布，拟合的程度一般由散度指标来度量。然而散度对于模型诊断所提供的信息有限，因此在模型研究或应用的评价阶段，研究者往往使用质量和多样性两类指标进行联合评价。尽管这种联合评价的方式已被广泛应用于新模型的评价，两者之间所呈现的互为权衡的现象仍没有可靠的解释，而且这种评价方式是否还能充分反映原来分布拟合的目标仍是一个未知的问题。本文主要针对以上问题展开理论和实验上的分析，已被 ICML2020 录用。



本文通过严谨的数学证明，得到以下两条主要结论：

- (1) 质量和多样性指标两者的确互为权衡。这为其他研究者广泛观察到的质量 - 多样性权衡现象提供了理论解释。
- (2) 当且仅当质量和多样性指标的形式满足一个严格的条件时，两者的线性组合可以构成一个散度指标。这意味着随意搭配质量和多样性的指标很可能无法反映分布拟合的程度。

本文从理论和实验上验证了广泛使用的质量 / 多样性指标组合 BLEU/Self-BLEU 无法准确反映分布拟合程度。为了解决以上问题，本文提出了新的指标组合，即覆盖率 (CR) 作为质量指标，负重复率 (NRR) 作为多样性指标。

$$\text{CR}_n(Q; P) = \sum_{gram_n \in S_n} Q_g(gram_n) \cdot P_g(gram_n)$$
$$\text{NRR}_n(Q) = - \sum_{gram_n \in S_n} Q_g^2(gram_n)$$

理论和实验均显示 CR/NRR 指标组合可以构成散度指标，且两者可以分别反映质量和多样性。

Beijing Academy of Artificial Intelligence



微信关注

北京智源人工智能研究院