

《消极评论的严重程度》

组员：甘指晴 2006500026

黄咏俊 2006500028

曹文彬 2006500027

kaggle 链接:

[Jigsaw Rate Severity of Toxic Comments | Kaggle](https://www.kaggle.com/competitions/jigsaw-toxic-severity-rating/overview)

<https://www.kaggle.com/competitions/jigsaw-toxic-severity-rating/overview>

科学意义:

当我们要求人类在没有任何上下文的情况下判别某个评论，以决定哪些是有毒的，哪些是无害的，这不是一件容易的事。此外，每个人都可能有自己的判断标准。我们试图通过汇总多数票的决定来解决这个问题。但许多研究人员正确地指出，这种做法会丢弃有意义的信息。一个简单得多的任务是问人们，他们觉得两种评论中哪一种更有害。但如果两种评论都是无害的，人们通常会随机选择。当一个评论看起来显然是正确的选择时，评判者之间的一致性结果要高得多。

科学问题:

在这次比赛中，研究人员将成对的评论呈现给专家打分，他们根据自己对毒性的看法，从两条评论中选出一条更有害。当你为评论提供分数时，应当把重点放在对评论毒性的严重程度进行排序，从无害到令人愤慨。其中，中间和极端一样重要。

数据集:

每条样本给定一对消极评论样例，并对其消极严重程度(毒性)进行评分，与毒性程度较低的评论相比，毒性程度较高的评论应获得更高的数值。本数据集共包含 30109 条样本。注：①本次比赛的数据集包含可能被视为亵渎、粗俗或冒犯的文本。②分数是相对的，不受特定值范围的约束

计划使用的算法：self-attention、Transformer、gradient descent、back propagation

预计达到的目标:

利用信息熵等概念实现对有毒有害评论的评分，从而相对准确地量化消极评论的毒性程度。

