

# C 题第一问补充

预测区间如何表示

对于 2023 年 3 月 1 日的预测值为  $a$  时, 误差为  $b, w$  则

$$\text{区间为 } \left[ \frac{a}{1+w}, \frac{a}{1-w} \right]$$

对于困难模型而言 令困难系数为  $B$ , 总数为  $DA$ , 则

$$f = \frac{B}{A} \quad \text{则 } f \text{ 表示困难模型占比}$$

若分别对 1~7 次打分为 6, 5, 4, ..., 0. 则

令每次 1~7 次的得分占比  $P_1, P_2, \dots, P_7$ .

$$\text{则 } \sum_{i=1}^7 P_i = 1$$

$$P = 6P_1 + 5P_2 + 4P_3 + 3P_4 + 2P_5 + P_6$$

$P$  表示平均分

分别按照单词属性对单词分类, 从而研究  $P$  与  $f$  的关系

以名词为例 通过对数据收集得到下表.

序号	困难系数	平均分 (平均分 $P$ )

1-①



通过回归拟合得到  $p$  与  $f$  的关系为:

$$p = g(f)$$

而  $g'(f)$  表示困难模型下得分增长率.

同理得到不同属性单词的函数关系

单词属性	$g(f)$	-阶导数
	$g_1(f)$	$g'_1(f)$
	$g_2(f)$	$g'_2(f)$
	$g_3(f)$	$g'_3(f)$

综上从而得到困难模型下单词属性不同对得分

的影响关系.

1-②

