TF-IDF

Web2.7

May 18, 2025

1 TF-IDF

1.1 TF

Term Frequency

$$TF(t,d) = \frac{n_{t,d}}{\sum_k n_{k,d}}$$

- $n_{t,d}$ t d
- $\sum_k n_{k,d}$ d
- : /

1.2 IDF

Inverse Document Frequency

$$IDF(t) = \log \frac{N}{df_t + 1}$$

- N
- df_t t
- 1 0
- : log(/ (+ 1))

1.3 TF-IDF

TF-IDF TF IDF

$$TFIDF(t,d) = TF(t,d) \times IDF(t)$$

- : (,) × ( )

2

2.1

$$Score_{title}(t) = TFIDF(t,d) \times (1 + \alpha_{title})$$

$\alpha_{title} = 0.8$

- : TF-IDF (,) × (1 +)

2.2

$$Score_{position}(t) = Score_{title}(t) \times (1 + \alpha_{position})$$

$\alpha_{position} = 0.5$

- **:** ( ) × (1 +)

2.3

$$Score_{tag} = \sum_{t \in keywords} Score_{position}(t) + \beta_{match} \times IsMatch_{title}(t)$$

- $keywords$
- $\beta_{match} = 0.7$
- $IsMatch_{title}(t)$
- **:** $\sum_{\in}$ ( ) + ×

3

3.1

$$Selected_{tag} = \begin{cases} 1 & \text{if } Score_{tag} \geq \theta \\ 0 & \text{otherwise} \end{cases}$$

$\theta = 0.005$

- **:** ≥ 1 ( ) 0

3.2

k $k = 5$

- **:** 5