

Similarity function in kBSPS kernel

The tolerant score function (t-score) is defined on two vertex-walks, t and s , of equal length q :

$$\text{t-score}_{L,E,D,l,e,d}(t,s) = \begin{cases} 0, & \text{if } (L = -1 \text{ and } \exists i : t[i] \neq s[i] \text{ and } \text{type}(t[i], s[i]) = l) \\ & \text{or } (E = -1 \text{ and } \exists i : t[i] \neq s[i] \text{ and } \text{type}(t[i], s[i]) = e) \\ & \text{or } (E = -1 \text{ and } \exists i : \text{type}(t[i]) \text{ or } \text{type}(s[i]) \text{ is entity and } \text{type}(t[i], s[i]) = 0) \\ & \text{or } (D = -1 \text{ and } \exists i : t[i] \neq s[i] \text{ and } \text{type}(t[i], s[i]) = d) \\ \sum_{i=1}^{2q-1} \text{type}(t[i], s[i]) \cdot \delta(t[i], s[i]), & \text{otherwise} \end{cases}$$

where $\text{type}(t[i]) = \{\text{surface,entity,dependency}\}$ is the type of $t[i]$, the i th element of the v-walk t ,

$$\text{type}(t[i], s[j]) = \begin{cases} l, & \text{if } \text{type}(t[i]) = \text{type}(s[j]) = \text{surface} \\ e, & \text{if } \text{type}(t[i]) = \text{type}(s[j]) = \text{entity} \\ d, & \text{if } \text{type}(t[i]) = \text{type}(s[j]) = \text{dependency} \\ 0, & \text{otherwise} \end{cases}$$

assigns the token-type-dependent weight if the tokens are of the same type, and

$$\delta(t[i], s[j]) = \begin{cases} 1, & \text{if } t[i] = s[j] \\ 0, & \text{otherwise} \end{cases}$$

is the Kronecker-delta defined on tokens. Figure 3 shows some examples for the t-score calculation.