Derivatives of RF

1 $BM25F$

$BM25F$ is defined by (1),(2) and (3):

$$BM25F(d,q) = \sum_{t \in d \cap q} IDF(t) \times \frac{TF_D(d,t)}{k_1 + TF_D(d,t)} \tag{1}$$

$$IDF(t) = log \frac{N}{N_d} \tag{2}$$

$$TF_D(d,t) = w_{summ} \times \frac{occurrences(d[summ],t)}{1 - b_{summ} + \frac{b_{summ} \times length_{summ}}{average\_length_{summ}}} + w_{desc} \times \frac{occurrences(d[desc],t)}{1 - b_{desc} + \frac{b_{desc} \times length_{desc}}{average\_length_{desc}}} \tag{3}$$

Derivative of $BM25F$ with respect to k_1:

$$\frac{\partial BM25F}{\partial k_1} = - \sum_{t \in d \cap q} IDF(t) \times \frac{TF_D(d,t)}{(k_1 + TF_D(d,t))^2} \tag{4}$$

Derivative of $BM25F$ with respect to w_{summ}:

$$\frac{\partial BM25F}{\partial w_{summ}} = \sum_{t \in d \cap q} IDF(t) \times \frac{k_1}{(k_1 + TF_D(d,t))^2} \times \frac{occurrences(d[summ],t)}{1 - b_{summ} + \frac{b_{summ} \times length_{summ}}{average\_length_{summ}}} \tag{5}$$

Derivative of $BM25F$ with respect to w_{desc}:

$$\frac{\partial BM25F}{\partial w_{desc}} = \sum_{t \in d \cap q} IDF(t) \times \frac{k_1}{(k_1 + TF_D(d,t))^2} \times \frac{occurrences(d[desc],t)}{1 - b_{desc} + \frac{b_{desc} \times length_{desc}}{average\_length_{desc}}} \tag{6}$$

Derivative of $BM25F$ with respect to b_{summ}:

$$\frac{\partial BM25F}{\partial b_{summ}} = \sum_{t \in d \cap q} IDF(t) \times \frac{k_1}{(k_1 + TF_D(d,t))^2} \times \frac{w_{summ} \times occurrences(d[summ],t)}{(1 - b_{summ} + \frac{b_{summ} \times length_{summ}}{average\_length_{summ}})^2} \times (1 - \frac{length_{summ}}{average\_length_{summ}}) \tag{7}$$

Derivative of $BM25F$ with respect to b_{desc}:

$$\frac{\partial BM25F}{\partial b_{desc}} = \sum_{t \in d \cap q} IDF(t) \times \frac{k_1}{(k_1 + TF_D(d,t))^2} \times \frac{w_{desc} \times occurrences(d[desc],t)}{(1 - b_{desc} + \frac{b_{desc} \times length_{desc}}{average\_length_{desc}})^2} \times (1 - \frac{length_{desc}}{average\_length_{desc}}) \tag{8}$$

2 $BM25F_{ext}$

$BM25F_{ext}$ is defined by (9) and (10):

$$BM25F(d,q) = \sum_{t \in d \cap q} IDF(t) \times \frac{TF_D(d,t)}{k_1 + TF_D(d,t)} \times \frac{(k_3 + 1) \times TF_Q(q,t)}{k_3 + TF_Q(q,t)} \tag{9}$$

$$TF_Q(q,t) = w_{summ} \times occurrences(d[summ],t) + w_{desc} \times occurrences(d[desc],t) \tag{10}$$

2.1 First Round

In the first tuning round, k_3 is fixed to 0, and $BM25F_{ext}$ is reduced to $BM25F$, thus

Derivative of $BM25F_{ext}$ with respect to k_1:

$$\frac{\partial BM25F_{ext}}{\partial k_1} = \frac{\partial BM25F}{\partial k_1} \tag{11}$$

Derivative of $BM25F_{ext}$ with respect to w_{summ}:

$$\frac{\partial BM25F_{ext}}{\partial w_{summ}} = \frac{\partial BM25F}{\partial w_{summ}} \tag{12}$$

Derivative of $BM25F_{ext}$ with respect to w_{desc}:

$$\frac{\partial BM25F_{ext}}{\partial w_{desc}} = \frac{\partial BM25F}{\partial w_{desc}} \tag{13}$$

Derivative of $BM25F_{ext}$ with respect to b_{summ}:

$$\frac{\partial BM25F_{ext}}{\partial b_{summ}} = \frac{\partial BM25F}{\partial b_{summ}} \tag{14}$$

Derivative of $BM25F_{ext}$ with respect to b_{desc}:

$$\frac{\partial BM25F_{ext}}{\partial b_{desc}} = \frac{\partial BM25F}{\partial b_{desc}} \tag{15}$$

2.2 Second Round

In the second round, all k_1, w_{summ}, w_{desc}, b_{summ} and b_{desc} are fixed except k_3, thus

Derivative of $BM25F_{ext}$ with respect to k_3:

$$\frac{\partial BM25F_{ext}}{\partial k_3} = \sum_{t \in d \cap q} IDF(t) \times \frac{TF_D(d,t)}{k_1 + TF_D(d,t)} \times \frac{{TF_Q}^2 - TF_Q}{(k_3 + TF_Q)^2} \tag{16}$$

3 *RF*

The retrieval function RF is defined in (17):

$$RF(d,q) = \sum_{i=1}^{7} w_i \times feature_i \tag{17}$$

Derivative of RF with respect to w_i:

$$\frac{\partial RF}{\partial w_i} = feature_i \tag{18}$$

Derivative of RF with respect to a parameter p in $feature_i (i \in \{1,2\})$ – the textual similarities computed by $BM25F_{ext}$:

$$\frac{\partial RF}{\partial p} = w_i \times \frac{\partial BM25F_{ext}}{\partial p} \tag{19}$$

4 RNC

Given the retrieval function RF and a training instance $I = (q, rel, irr)$, the cost function RNC is defined in (20):

$$RNC(I) = log(1 + e^Y) \text{ where } Y = RF(irr,q) - RF(rel,q) \tag{20}$$

Derivative of RNC with respect to Y:

$$\frac{\partial RNC}{\partial Y} = \frac{e^Y}{1 + e^Y} \tag{21}$$

Derivative of Y with respect to a parameter x in RF:

$$\frac{\partial Y}{\partial x} = \frac{\partial RF}{\partial x}(irr,q) - \frac{\partial RF}{\partial x}(rel,q) \tag{22}$$

Therefore, the derivative of RNC with respect to a parameter x in RF is defined as follows:

$$\frac{\partial RNC}{\partial x} = \frac{\partial RNC}{\partial Y} \times \frac{\partial Y}{\partial x} = \frac{\partial RNC}{\partial Y} \times (\frac{\partial RF}{\partial x}(irr,q) - \frac{\partial RF}{\partial x}(rel,q)) \tag{23}$$