Convergence Proof

Let $V = [V_{ij}] \in \mathbb{R}^{q\times p}$ be the initial relevant score matrix with candidate documents $\{D_{1,1},\ldots,D_{k,t(k)}\}$ and systems $\{S_1,\ldots,S_p\}$, where $q=\sum_{i=1}^{k} t(k)$. The EM-based pseudo-judgment estimating process can be simply described as the operations on the score matrix V and two vectors: the weights of systems $w = (w_1,...,w_p)^T$ and the pseudo-judgments $J=(J_{1,1},...,J_{k,t(k)})^T$. According to the definition of pseudo-judgment score in the paper:

$$J^t_{m,n} = \sum_{j=1}^{p} w^t_j \cdot f(V_{S_j,D_{m,n}}) \tag{1}$$

The pseudo-judgments at the t-step J^t can be represented as:

$$J^t = (J^t_{1,1},...,J^t_{k,t(k)})^T = Vw^t \tag{2}$$

where $w^t = (w^t_1,...,w^t_p)^T$ is the weights at t-step. For system S_j, the relevant scores v_j can be represented as:

$$v_j = (V_{s_j,D_{1,1}},...,V_{s_j,D_{k,t(k)}})^T = Ve_j \tag{3}$$

with $\mathbf{e}_j = (\underbrace{0,\ldots,0}_{j-1},1,0,\ldots,0)^T \in \mathbb{R}^p$.

With the previous definition, the algorithm can be described in the following four steps:

- Step 1: Let $w^1 = (w^1_1,\ldots,w^1_p)^T = (\frac{1}{p},...,\frac{1}{p})$ be the initial weights for the systems $\{S_1,\ldots,S_p\}$ at 0-step.
- Step 2: According to Equation (1), the pseudo-judgments for the candidate documents at t-step will be $J^t = Vw^t$. Then define the loss values $L^t := (L^t_1,\ldots,L^t_p)^T$ with the j-th loss value L^t_j for system S_j as:

$$L^t_j = (Ve_j - Vw^t)^T(Ve_j - Vw^t) \tag{4}$$
$$= (e_j - w^t)^T V^T V(e_j - w^t) \tag{5}$$

- Step 3: Obtain the new $w^{t+1} = (w^{t+1}_1,\ldots,w^{t+1}_p)^T$ for $(t+1)$-step with system s_j's weight w^{t+1}_j

$$w^{t+1}_j := \frac{C_t - L^t_j}{\sum_{j=1}^{p}(C_t - L^t_j)} \tag{6}$$

  where

$$C_t := \|Vw^t\|_2^2 + 1 \tag{7}$$

– Step 4: Repeat step 2 and step 3 until the process converges. The final vector J will be the real-valued estimating scores for the pseudo-judgments.

Theorem 1. *The algorithm converges in exponential rate.*

Proof. From Equations (5) and (6), we have

$$w_j^{t+1} = \frac{C_t - (e_j - w^t)^T V^T V (e_j - w^t)}{\sum_{j=1}^{p}(C_t - (e_j - w^t)^T V^T V (e_j - w^t))} \tag{8}$$

$$= \frac{C_t - e_j^T V^T V e_j - (w^t)^T V^T V w^t + 2e_j^T V^T V w^t}{pC_t - \sum_{j=1}^{p}(e_j^T V^T V e_j + (w^t)^T V^T V w^t - 2e_j^T V^T V w^t)} \tag{9}$$

Accordingly, we have

$$w_j^{t+1} = \frac{C_t - \|Vw^t\|_2^2 - v_j^T v_j + 2v_j^T V w^t}{p(C_t - \|Vw^t\|_2^2) - \sum_{j=1}^{p} v_j^T v_j + 2\sum_{j=1}^{p} v_j^T V w^t} \tag{10}$$

In which v_j is defined in Equation (2). Rescale each column of V such that $v_j^T v_j = 1$ for $(j = 1, \ldots, p)$. We have

$$w_j^{t+1} = \frac{C_t - \|Vw^t\|_2^2 - 1 + 2v_j^T V w^t}{p(C_t - \|Vw^t\|_2^2 - 1) + 2\sum_{j=1}^{p} v_j^T V w^t} \tag{11}$$

Recalling that $C_t := \|Vw^t\|_2^2 + 1$, we have

$$w_j^{t+1} = \frac{v_j^T V w^t}{\sum_{j=1}^{p} v_j^T V w^t} = \frac{(V^T V w^t)_j}{\|V^T V w^t\|_1} \tag{12}$$

$$\Rightarrow \; w^{t+1} = \frac{V^T V w^t}{\|V^T V w^t\|_1} \tag{13}$$

Denote by $M = V^T V$, we have

$$w^{t+1} = \frac{Mw^t}{\|Mw^t\|_1} = \frac{M^{t+1} w^1}{\|M^{t+1} w^1\|_1} \tag{14}$$

This is similar to the power iteration and the convergence is geometric. In detail, the power iteration algorithm starts with a vector x_0. The method is described by the iteration

$$x_{t+1} = \frac{\mathbf{A} x_t}{\|\mathbf{A} x_t\|_1} \tag{15}$$

Under the assumptions:

– **A** has an eigenvalue that is strictly greater in magnitude than its other eigenvalues.

– The starting vector x_0 has a nonzero component in the direction of an eigenvector associated with the dominant eigenvalue.

Then the subsequence of x_t will converge to an eigenvector associated with the dominant eigenvalue with convergence ratio $\frac{|\lambda_1|}{|\lambda_2|}$, where λ_1 is the dominant eigenvalue and λ_2 is the second dominant eigenvalue.