Convergence Proof

Let $V = [V_{ij}] \in \mathbb{R}^{q \times p}$ be the initial relevant score matrix with candidate documents $\{D_{1}, \ldots, D_{k(k)}\}$ 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 $w = (w_{1}, \ldots, w_{p})^{T}$ and the pseudo-judgments $J = (J_{1}, \ldots, J_{k(k)})^{T}$. According to the definition of pseudo-judgment score in the paper:

$$J'_{m,n} = \sum_{j=1}^{p} w'_{j} \cdot f(V_{j}, D_{m,n})$$  \hspace{1cm} (1)

The pseudo-judgments at the $t$-step $J'$ can be represented as:

$$J' = (J'_{1}, \ldots, J'_{k(k)})^{T} = Vw'$$  \hspace{1cm} (2)

where $w' = (w'_{1}, \ldots, w'_{p})^{T}$ is the weights at $t$-step. For system $S_{j}$, the relevant scores $v_{j}$ can be represented as:

$$v_{j} = (V_{s_{j}, D_{1}, 1}, \ldots, V_{s_{j}, D_{k(t(k)}})^{T} = Ve_{j}$$  \hspace{1cm} (3)

with $e_{j} = (0, \ldots, 0, 1, 0, \ldots, 0)^{T} \in \mathbb{R}^{p}$.

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

- Step 1: Let $w^{1} = (w^{1}_{1}, \ldots, w^{1}_{p})^{T} = (\frac{1}{p}, \ldots, \frac{1}{p})$ 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 $J' = Vw'$. Then define the loss values $L' := (L'_{1}, \ldots, L'_{p})^{T}$ with the $j$-th loss value $L'_{j}$ for system $S_{j}$ as:

$$L'_{j} = (Ve_{j} - Vw')^{T} (Ve_{j} - Vw')$$  \hspace{1cm} (4)

$$= (e_{j} - w')^{T} V^{T} V (e_{j} - w')$$  \hspace{1cm} (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 $w^{t+1}_{j}$

$$w^{t+1}_{j} := \frac{C_{t} - L'_{j}}{\sum_{j=1}^{p} (C_{t} - L'_{j})}$$  \hspace{1cm} (6)

where

$$C_{t} := ||Vw'||_{2}^{2} + 1$$  \hspace{1cm} (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

\[
\begin{align*}
  w_{j}^{t+1} &= \frac{C_{t} - (e_{j} - w_{t}^{t})V^{T}V(e_{j} - w_{t}^{t})}{\sum_{j=1}^{p}(C_{t} - (e_{j} - w_{t}^{t})V^{T}V(e_{j} - w_{t}^{t}))} \\
  &= \frac{C_{t} - e_{j}^{T}V^{T}V e_{j} - (w_{t}^{t})^{T}V^{T}V w_{t}^{t} + 2e_{j}^{T}V^{T}V w_{t}^{t}}{pC_{t} - \sum_{j=1}^{p}(e_{j}^{T}V^{T}V e_{j} + (w_{t}^{t})^{T}V^{T}V w_{t}^{t} - 2e_{j}^{T}V^{T}V w_{t}^{t})} \\
\end{align*}
\]

(8)

Accordingly, we have

\[
\begin{align*}
  w_{j}^{t+1} &= \frac{C_{t} - \|V w_{t}^{t}\|_{2}^{2} - v_{j}^{T}V w_{t}^{t} + 2v_{j}^{T}V w_{t}^{t}}{p(C_{t} - \|V w_{t}^{t}\|_{2}^{2} - 1) + 2\sum_{j=1}^{p}v_{j}^{T}V w_{t}^{t}} \\
\end{align*}
\]

(9)

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

\[
\begin{align*}
  w_{j}^{t+1} &= \frac{C_{t} - \|V w_{t}^{t}\|_{2}^{2} - 1 + 2v_{j}^{T}V w_{t}^{t}}{p(C_{t} - \|V w_{t}^{t}\|_{2}^{2} - 1) + 2\sum_{j=1}^{p}v_{j}^{T}V w_{t}^{t}} \\
\end{align*}
\]

(10)

Recalling that \( C_{t} := \|V w_{t}^{t}\|_{2}^{2} + 1 \), we have

\[
\begin{align*}
  w_{j}^{t+1} &= \frac{v_{j}^{T}V w_{t}^{t}}{\sum_{j=1}^{p}v_{j}^{T}V w_{t}^{t}} = \frac{(V^{T}V w_{t}^{t})_{j}}{\|V^{T}V w_{t}^{t}\|_{1}} \\
  \Rightarrow w_{j}^{t+1} &= \frac{V^{T}V w_{t}^{t}}{\|V^{T}V w_{t}^{t}\|_{1}} \\
\end{align*}
\]

(12)

(13)

Denote by \( M = V^{T}V \), we have

\[
\begin{align*}
  w_{t+1} &= \frac{M w_{t}}{\|M w_{t}\|_{1}} = \frac{M^{t+1} w_{1}}{\|M^{t+1} w_{1}\|_{1}} \\
\end{align*}
\]

(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

\[
\begin{align*}
  x_{t+1} &= \frac{Ax_{t}}{\|Ax_{t}\|_{1}} \\
\end{align*}
\]

(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 $\lambda_1$ is the dominant eigenvalue and $\lambda_2$ is the second dominant eigenvalue.