



iii.

Apply SeNMF with k latent topics to each member of the TF-IDF and SSPMI ensembles, \mathcal{X}_q , \mathcal{M}_q $W_q, \mathcal{H}_q, \mathcal{G}_q = \min D_{KL}(\mathcal{X}_q || WH) + \alpha D_{KL}(\mathcal{M}_q || WG)$ subject to $W \ge 0, H \ge 0, G \ge 0$ for $1 \le q \le r$ Cluster the set of extracted topics, $\{\boldsymbol{\mathcal{W}}_1 \ \boldsymbol{\mathcal{W}}_2 \ \dots \ \boldsymbol{\mathcal{W}}_r\}$ into k clusters and compute mediods, \tilde{W} . Compute silhouette statistics Minimize $D_{KL}(\boldsymbol{X}||\tilde{\boldsymbol{W}}\boldsymbol{H})$ on clustered $\{\boldsymbol{\mathcal{W}}_1 \ \boldsymbol{\mathcal{W}}_2 \ \dots \ \boldsymbol{\mathcal{W}}_r\}$. to find $ilde{m{H}}$ Determine correct latent dimension, k, using L-statistics, and cluster columns of \tilde{H} to identify the groups of similar documents in text corpus.

İ۷.