

1 Summary

Structural re-ranking to improve the precision on the highest ranked list ranks. The general idea is that the ranked list can be represented as a *one-way bipartite graph* of individual documents and their clusters. Following types of graphs are considered:

- **document-to-document**
- **cluster-to-document**
- **document-to-cluster.**

Hypothesis 1: Document centrality can be used for determining the re-ranking scheme.

Hypothesis 1 assumes that the initial retrieved sets (in this work, top 50 documents) already contains good selection of relevant document, but their rankings can be improved by placing the documents that are related to many other documents in the set higher (*central documents*). HITS and PageRank are used to determine such central documents.

2 Background

- Graph edges weights $w_{u \rightarrow v}$ in a non-hypertext environment are defined as

$$w_{u \rightarrow v} = e^{-KL(u||v)}.$$

- HITS algorithm has a recursive definition

$$auth(v) = \sum_{u \in V} w_{u \rightarrow v} hub(u)$$

$$hub(u) = \sum_{v \in V} w_{u \rightarrow v} auth(v),$$

which converges for one-way bipartite graphs.

- PR can be defined for one-way bipartite graphs as

$$PR_{bip}(v) = \sum_{u \in V} \frac{w_{u \rightarrow v}}{out(u)}.$$

3 Re-Ranking algorithms

1. **PageRank. $D2D$ graph.**
 - (a) Rank documents according to their PR only.
 - (b) Rank documents according to $PR * p(q|d)$.
2. **HITS. $D2D$ graph.** Rank documents according to their authority score.
3. **HITS. $C2D$ graph.** Clusters are hubs, documents are authorities. Rank documents according to their authority score.
4. **HITS. $D2C$ graph.** Documents are hubs, clusters are authorities. Rank clusters according to their authority score. Then, rank documents inside each cluster according to $p(q|d)$.

4 Experiments

- $PR * p(q|d)$ is significantly better than PR . (Result from [1])
- $HITS - C2D$ is better than both $PR - D2D$ and $HITS - D2D$. *Clusters are beneficial for document re-ranking.* (Table 1.)
- $HITS - D2C$ is better in identifying clusters that contain a high percentage of relevant documents than cluster language model method discussed in [2]. (Table 3.)

5 Related Work

1. O. Kurland and L. Lee. "Page rank without hyperlinks."
2. X.Liu and W. B. Croft. "Cluster-based retrieval using language models."