Eliminating Redundancy by Spectral Relaxation for Multi-Document Summarization

F. Fukumoto, A. Sakai and Y. Suzuki

Univ. of Yamanashi

Graph-based multi-document summarization

- LexPagerank (Erkan and Radev,2004)
- PageRank and HITS (Mihalcea and Tarau,2005)

  - Constructing graph consisting nodes and links
  - Applying graph-based ranking algorithm
  - Chose the sentences with large rank score into the summary

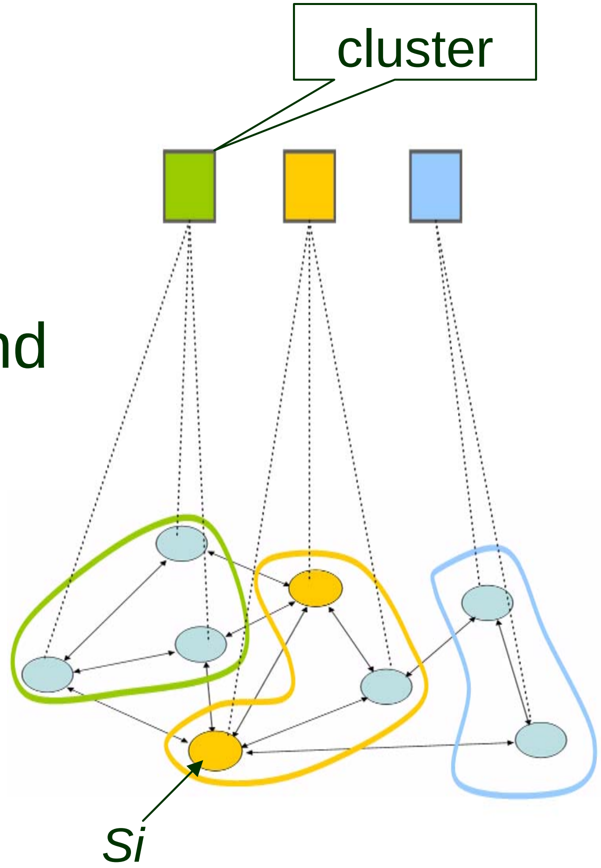
All the sentences are ranked based on a sentence as unit of information.

Semantically related two sentences with “high recommendation” are ranked with high score, and thus are regarded as a summary sentence.

The resulting summary still contains overlapping info.

Cluster-based Multi-document Summarization

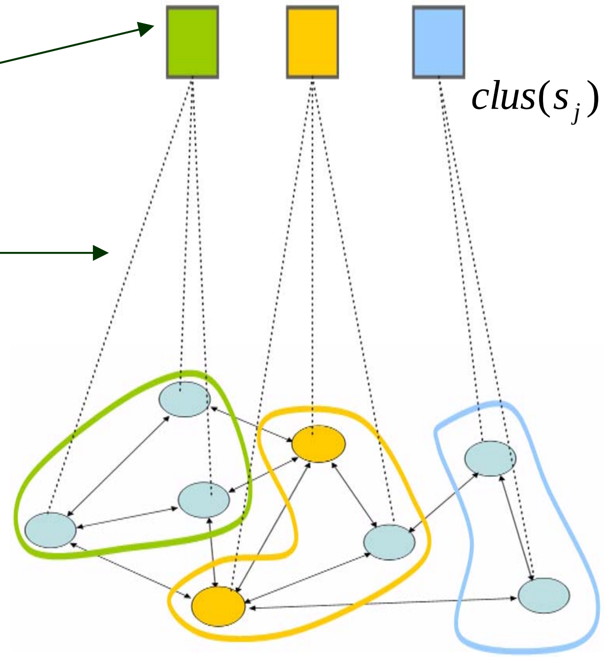
- ClusterCMRW model (X.Wan et al)
  - Classifying documents into theme clusters by using k-means
  - Constructing a graph to reflect the relationships between sentences and clusters by using MRW model
- Spectral Clustering (Weiss et al)
  - A transformation of the original sentences into a set of orthogonal eigenvectors.



Sentence extraction by ClusterCMRW

1. Weight between two sentences, conditioned on the two clusters containing the two sentences.

$$f(i \to j \mid clus(s_i), clus(s_j)) = f(i \to j) \cdot \{\lambda \cdot \boxed{\pi(clus(s_i))} \cdot \boxed{\omega(s_i, clus(s_i))} + (1-\lambda) \cdot \pi(clus(s_j)) \cdot \omega(s_j, clus(s_j))\}$$



$$\pi(clus(s_i)) = sim_{\cos ine}(clus(s_i), D)$$

$$\omega(s_i, clus(s_i)) = sim_{\cos ine}(s_i, clus(s_i))$$

$\pi(clus(s_i))$: The importance of the cluster $clus(s_i)$ in the document set *D*

$\omega(s_i, clus(s_i))$: The correlation between the sentence s_i and its cluster $clus(s_i)$

Sentence extraction by ClusterCMRW

2. The transition probability from s_i to s_j

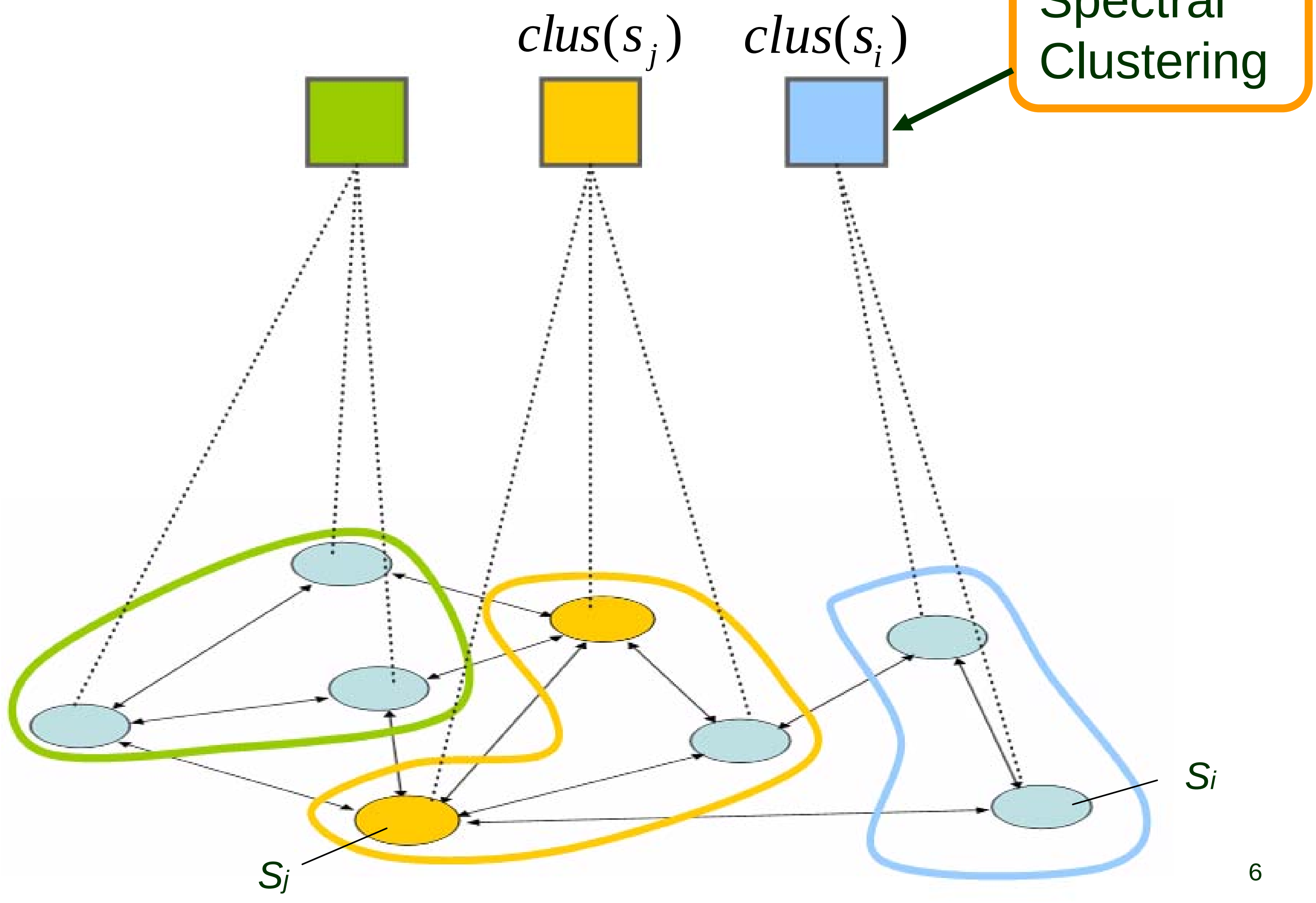
$$p(i \to j \mid clus(s_i), clus(s_j)) = \frac{f(i \to j \mid clus(s_i), clus(s_j))}{\sum_{k=1}^{|V|} f(i \to k \mid clus(s_i), clus(s_j))}$$

3. The final transition matrix $\tilde{M}_{ij}$

$$\tilde{M}_{ij} = p(i \to j \mid clus(s_i), clus(s_j))$$

$$\vec{\lambda}^{*} = \mu \tilde{M}^{*T} + \frac{(1-\mu)}{|V|} \vec{e}\,\vec{e}^{T}$$

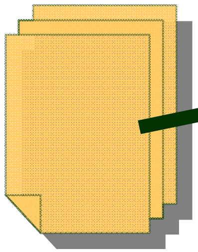
ClusterCMRW model



Sentence Classification by Spectral Clustering

1. Form a distance matrix *D*

Multi-doc.



Word freq.

$$\begin{array}{c|c} & \begin{array}{cccc} w1 & w2 & w3 & w4 \end{array} \\ \begin{array}{c} S1 \\ S2 \\ S3 \end{array} & \begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 \end{pmatrix} \end{array}$$

Euclid distance

$$D = \begin{bmatrix} 0 & 1 & 1.4 \\ 1 & 0 & 1 \\ 1.4 & 1 & 0 \end{bmatrix}$$

2. Feature space and sentence classification

● *D* is transformed to an affinity matrix *Aij*

$$A_{ij} = \begin{cases} \exp\left(-\dfrac{D_{ij}^2}{2\sigma^2} \right) & \text{if} \quad i \neq j \\ 0 & \text{if} \quad i = j \end{cases}$$

● Create a diagonal matrix *B*

$$B_{ii} = \sum_{j=1}^{n} A_{ij}$$

● Create L

$$L = B^{-1/2} A B^{-1/2}$$

- **Each item has a vector of l coordinates in the transformed space.**
- **These vectors are normalized to unit length, and K-means is applied to S in l-dimensional space.**

Multi-document summarization by ClusterCMRW

Experiments

1. Data

Source data	NTCIR-3 SUMM FBFREE DryRun and FormalRun (1998-1999 Japanese newspapers)
# of topics	**30**
# of sentences / doc	30 to 350 sentences
Ext. of sentences	NTCIR-3 SUMM FBFREE Long and short according to the character length
# of clusters	The square root of the number of sentences

Experiments

2. Two evaluation measures:

- Cosine similarity between the generated summary by the system and the human generated summary
- ROUGE score used in DUC

$$ROUGE = \frac{\sum_{S \in \{RefSum\}} \sum_{word \in S} Count_{match}(word)}{\sum_{S \in \{RefSum\}} \sum_{word \in S} Count(word)}$$

Parameter estimation used in the spectral clustering

- 10 topics to estimate two parameters σ and l in the l-dimensional space
  - σ is searched in steps of 0.01 from 1.0 to 5.0
  - l is searched in steps 10% from 0 to 80% against the total number of words in the training data

  - The size that optimized the average F-score of 10 topics was chosen
    - σ is set to 4.5
    - l is set to 80%

Summarization Results

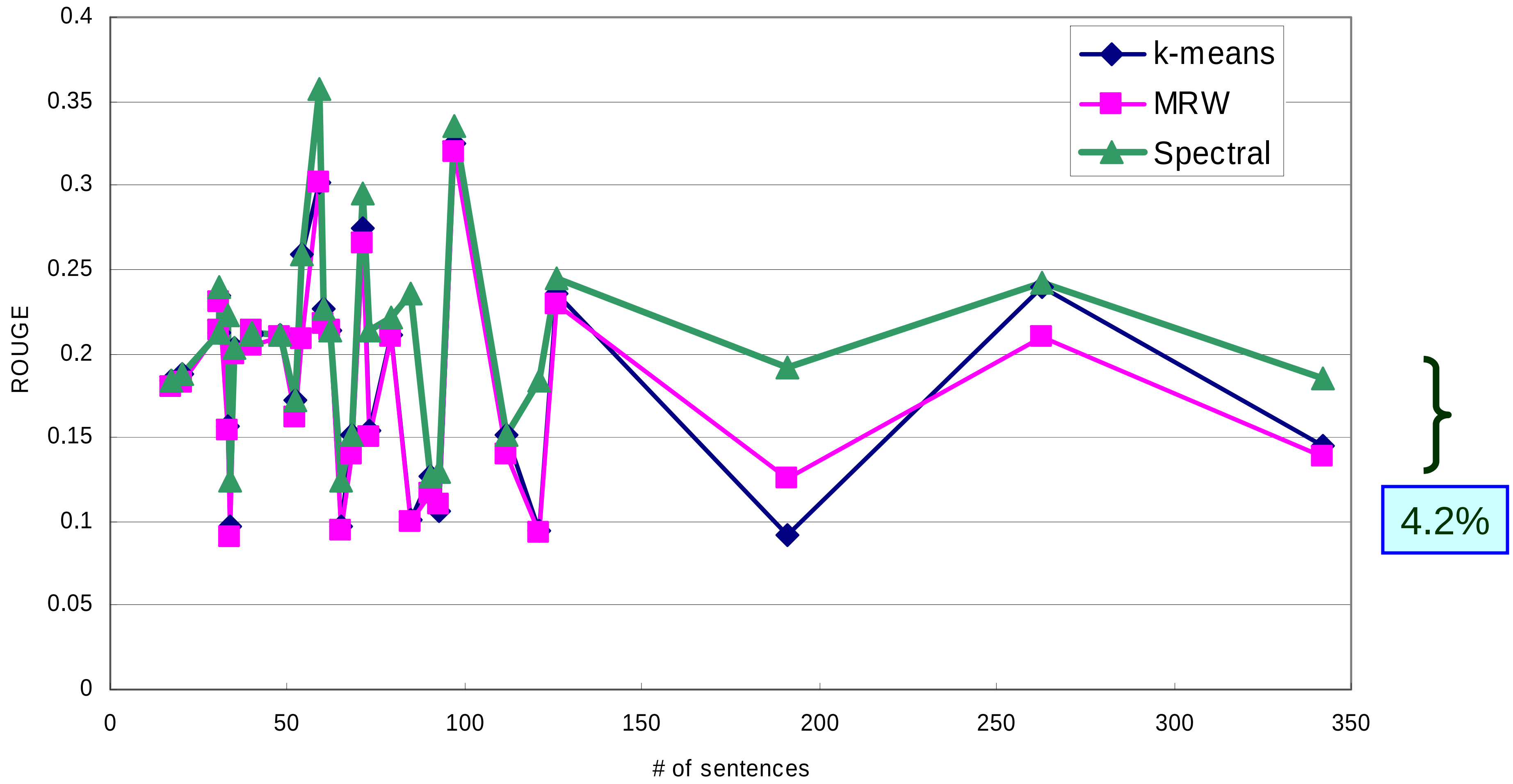
	# of doc	# of sent	# of sum	cos			ROUGE		
				MRW	K-means	Sp	MRW	K-means	Sp
short	7.5	83.0	11.9	0.431	0.575	0.632	0.330	0.334	0.360
long			20.4	0.371	0.408	0.477	0.180	0.186	0.209

- Sp outperformed the baselines, MRW and k-means, regardless of the types of summary, and evaluation measures
- Short was better than long. The rank score of correct sentences within the candidate sentences obtained by the MRW model works well.

Sentence Similarities within a summary

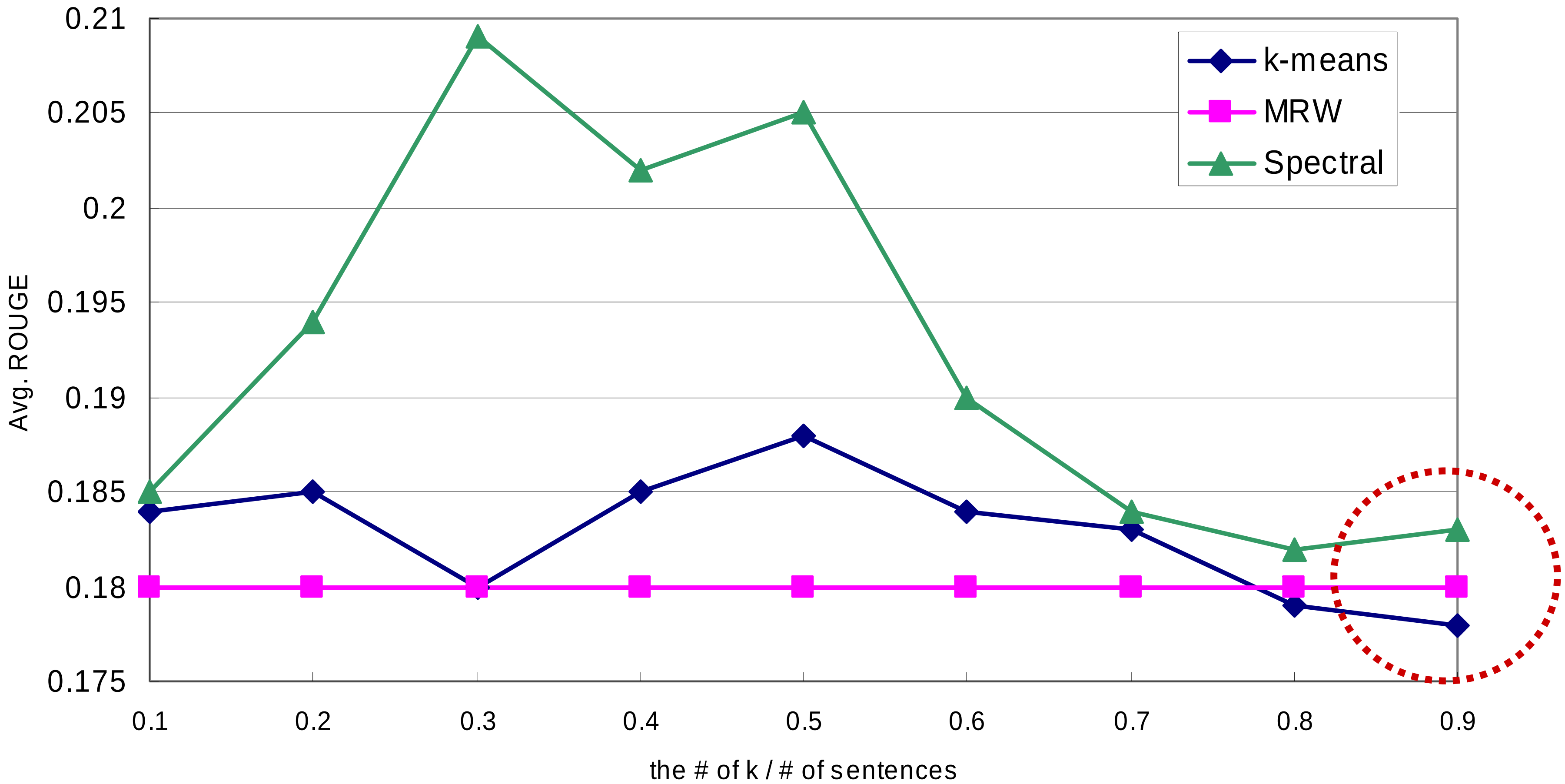
	# of doc	# of sent	# of sum	Similarity within a summary sentences			
				Human	MRW	K-means	Sp
short	7.5	83.0	11.9	0.129	0.137	0.142	0.132
long			20.4	0.173	0.283	0.201	0.193

of sentences vs ROUGE score



- SP is more robust than k-means and simple MRW model even for a large number of input sentences

of k vs ROUGE score



- Sp outperformed the results obtained by directly applying MRW.
- The results by k-means was worse than the results of MRW when the ratio of the # of cluster *k* against the # of sentences as an input was larger than 80%. For a large number of topics, k-means is not effective.

Conclusion

- A method to detect salient sentences from documents that discuss the same event
- 10.6% improvement over a baseline MRW (cosine), and 2.9% (ROUGE score)

- Applying the method to the DUC evaluation data
- Extending the method to classify sentences into more than one clusters by using soft-clustering techniques